It's Always Darkest Before the Cosmic Dawn:
Early Results from Novel Tools and Telescopes for
21 cm Cosmology

by

Joshua Shane Dillon

B.S., Stanford University (2009)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Author . . . .
Department of Physics
April 17, 2015

Certified by . . . .
Max Tegmark
Professor of Physics
Thesis Supervisor

Accepted by . . . .
Nergis Mavalvala
Professor of Physics
Associate Department Head for Education

It's Always Darkest Before the Cosmic Dawn:
Early Results from Novel Tools and Telescopes for
21 cm Cosmology

by

Joshua Shane Dillon

Abstract

21 cm cosmology, the statistical observation of the high redshift universe using the hyperfine transition of neutral hydrogen, has the potential to revolutionize our understanding of cosmology and the astrophysical processes that underlie the formation of the first stars, galaxies, and black holes during the "Cosmic Dawn." By making tomographic maps with low frequency radio interferometers, we can study the evolution of the 21 cm signal with time and spatial scale and use it to understand the density, temperature, and ionization evolution of the intergalactic medium over this dramatic period in the history of the universe.

For my Ph.D. thesis, I explore a number of advancements toward detecting and characterizing the 21 cm signal from the Cosmic Dawn, especially during its final stage, the epoch of reionization. In seven different previously published or currently submitted papers, I explore new techniques for the statistical analysis of interferometric measurements, apply them to data from current generation telescopes like the Murchison Widefield Array, and look forward to what we might measure with the next generation of 21 cm observatories. I focus in particular on estimating the power spectrum of 21 cm brightness temperature fluctuations in the presence enormous astrophysical foregrounds and how those measurements may constrain the physics of the Cosmic Dawn.

Thesis Supervisor: Max Tegmark
Title: Professor of Physics

Thus the explorations of space end on a note of uncertainty. And necessarily so. We are, by definition, at the very center of the observable region. We know our immediate neighborhood rather intimately. With increasing distance, our knowledge fades, and fades rapidly. Eventually, we reach the dim boundary—the utmost limits of our telescopes. There, we measure shadows, and we search among ghostly errors of measurement for landmarks that are scarcely more substantial.

The search will continue. Not until the empirical resources are exhausted, need we pass onto the dreamy realms of speculation.

EDWIN HUBBLE
The Realm of the Nebulae, 1936

Contents

Part I   Novel Data Analysis Tools   66

2 A Fast Method for Power Spectrum and Foreground Analysis for 21 cm Cosmology   67

Part II   Early Results from New Telescopes   **201**

Part III   The Cosmic Dawn on the Horizon   349

7   Detecting the 21 cm Forest in the 21 cm Power Spectrum   351

List of Figures

List of Tables

Acknowledgments

These past six months have put me in a reflective mood. Applying for postdocs and then writing this thesis have both been exercises in looking back, summarizing my graduate career, and then trying to project that momentum forward. As I come into my own as a physicist and move into this next stage of my career, I see that my time at MIT has given me a real running start. It's been a exhilarating and joyous six years and I have a lot of people to thank for that.

First, I have to thank my advisor, Max Tegmark. Max is responsible for guiding me on the difficult road from being interested in science to being a scientist. Of course, I learned a tremendous amount from him about how to tackle problems in cosmology and data analysis by dividing them into manageable parts with simple sense-checks. Max extended to me considerable freedom in working on what I found interesting while never letting me lose sight of the big picture. But more importantly, Max was, and I'm sure will continue to be, an invaluable source of sage career advice. I learned from Max how to pick interesting problems and good collaborators to work on them with and the essential importance of being nice. Max knew how important it was for me to work on HERA and I owe my postdoc in large part to his prescience.

Next I'd like to thank Jackie Hewitt. While not technically my advisor, I've worked closely enough with her over the last few years that she's basically been a co-advisor. Working with Jackie has afforded me the exciting and edifying opportunity to see how my theoretical techniques stand up to real data. Like Max, she always kept me focused on the big picture, though her hard-nosed realism always felt like a excellent counterpoint to Max's ebullience. Jackie is a model for leadership in academia and I'm grateful to have witnessed and benefited from her direction of both her research

group and the MKI.

The nexus of 21 cm cosmology that Max and Jackie have built at MIT has led to the incredible opportunity to collaborate with seven other graduate students all working toward the same goals. First and foremost, I have to thank Adrian Liu for serving as a mentor, role model, and mini-advisor. His perspective from a few years down basically the same road I'm traveling has been incredibly helpful. Adrian's selflessness as a collaborator is the model I have tried mimic (however imperfectly) in my own relationships with the younger 21 cm grad students, Aaron Ewall-Wice, Abraham Neben, and Jeff Zheng. I want to thank them for their tireless work and for the trust they placed in me to act as a mentor and sounding board, even if I sometimes led them down blind alleys. And of course, Chris Williams, Mike Matejek, and Andy Lutomirski have been patient tutors in all things radio astronomy and generous collaborators. The science I've done together with all seven friends and colleagues has been far more interesting, enjoyable, and fruitful than anything I could have done alone.

I have also had the great fortune to work with a number of outstanding collaborators in both the MWA and HERA. In particular, I am grateful for my recommenders, Aaron Parsons and Miguel Morales. I've already learned so much from both and them and I have no doubt that will continue as we build HERA togther. I also want to thank Jonnie Pober for his wonderful pessimism (and for letting me use his postdoc applications as a model), Danny Jacobs and Bryna Hazelton for reminding me that experimental reality is always more complicated than I think, Cath Trott for making me think deeply about estimators, and Adam Beardsley for the postdoc application commiseration.

Beyond my immediate research area, I am immensely grateful for the amazing intellectual community at MIT, especially in the astrophysics division. I found great friends, role models, and tutors among the older graduate students, including Ben Cain, Robyn Sanderson, Phil Zukin, Leo Stein, Leslie Rodgers, Scott Hertel, Nick Smith, John Rutherford, Uchupol Ruangsri, and Becky Levinson. I am honored and grateful to have shared the long road with my astrograd peers and dear friends—Kat

Sathirapongsasuti, Walter Vulej, Jon Kass Ian Counts, and Frank Wang. The Phi Psis really showed me around Boston and made me feel at home here and I am very grateful to have had an instant group of close friends when I moved here. Lastly, I am immeasurably thankful to have maintained such a close relationship with my high school friends (and, in a strange twist of fate, business partners) and I have to thank Daniel Dranove, Ben Hantoot, Eli Halpern, David Munk, David Pinsof, Max Temkin, and Eliot Weinstein for their years of true and loyal friendship that significantly predate this thesis and I'm sure will endure for many years beyond it.

Finally I need to thank my family. I am proud to come from a family that holds academic and intellectual pursuits in such high esteem. My parents and grandparents furnished me with every possible opportunity to discover my passions, pursue them, and excel in them. Even when there wasn't always money for the newest toy or a fancy vacation, there was always money for books. They taught both Molly and I the virtue of intellectual work in the service of humanity and I am grateful that we have encouraged each other in our individual pursuits of that goal. I can draw a straight line from those values to the completion of this thesis and my Ph.D. Their support, their pride in my accomplishments, and their love sustained this effort and for that I will always be grateful.

Preface

The Golden Age of Cosmology

I've often heard it said that we are living in a "golden age of cosmology." Perhaps just as often, I hear the equally sweeping claim that cosmology just finished its golden age and that all the exciting discoveries have probably been made. As I step back to survey the scientific landscape that I am graduating into—one that I hope to shape with my work—I have to ask myself: which is it?

The late University of Chicago cosmologist David Schramm is credited with first declaring the end of the twentieth century a golden age. In a meeting report on dark matter [195] he began:

> Let me open by noting that we're in the golden age of cosmology... Now cosmologists finally have the technology that allows experiments that tell us about the universe as a whole. We have been able to study it in a truly quantitative way, and we've been able to establish that the early universe was hot and dense.

By that he meant that the surprising discovery of the recession of almost all observed galaxies, coupled with the discovery for the cosmic microwave background and the precise measurement of the cosmic abundance of light elements, all upheld the remarkable theory that the universe began with a hot big bang. The weight of evidence had just reached the point where a basic framework could be worked out and (more or less) agreed upon—now it was time to fill in the gaps.

His sentiment was met with a mix of bemusement and skepticism. In one anecdote:

> He kept proclaiming that cosmology was in a "golden age." His chamber of commerce enthusiasm seemed to grate on some of his colleagues; after

all, one does not become a cosmologist to fill in the details left by pioneers. After Schramm's umpteeth "golden age" proclamation, one physicist snapped that you cannot know an age is golden when you are *in* that age but only in retrospect. Schramm jokes proliferated. One colleague speculated that the stocky physicist might represent the solution to the dark matter problem. Another proposed that Schramm be employed as a plug to prevent our universe from being sucked down a wormhole.

That story comes from John Horgan's provocatively titled *The End of Science* [96]. In it he argues that scientists across virtually all disciplines are already beginning to sense an end, a butting up against the limits of knowledge that comes with the extraordinary successes of fundamental science over the last few centuries. He worries that the "great revelations or revolutions" are behind us that that the ultimate *telos* of science—the "primordial quest to understand the universe and our place in it"—has been mostly accomplished. Writing on cosmology specifically, he asks:

> What if Schramm was right? What if cosmologists had, in the big bang theory, the major answer to the puzzle of the universe? What if all that remained was tying up loose ends, those that could be tied up?

I think Horgan misses the point. Schramm didn't think he lived in the golden age of cosmology because the biggest discoveries had just been made. It was a golden age because the recent triumph of the big bang model had opened up whole new lines of inquiry. Rather suddenly, cosmologists realized that they were solving an entirely different puzzle than they had been before. That doesn't mean that all pieces were in hand, or that all the pieces they'd find would fit in so neatly.

Thomas Kuhn, the philosopher and historian of science, famously wrote about this process in *The Structure of Scientific Revolutions* [112]. He describes the typical progress of an scientific discipline as puzzle solving or "normal science." Working within a shared framework, a community of scientists has common set of values and theories—a paradigm—which makes sensible a new set of questions about nature, a new set of puzzles to solve. When enough puzzles arise that linger unsolved as anomalies, a need arises for a new paradigm. Ideally it is more accurate, more predictive, of greater scope, and simpler than previous theories. Rarely is it so clear. In time, the

better theory wins the consensus, if perhaps not unanimous support. As Max Planck put it, "science advances one funeral at a time."

Our stories about science invariably romanticize the revolutionaries. We aspiring scientists all want to be the revolutionaries, but only a lucky few get the privilege. The more I study science, both its past and its present, the more I love the puzzle solving. I know that scientific revolution is impossible without puzzles that defy resolution, without the hard work that extracts from the full complexity of nature a slow trickle of anomalies.

This thesis is about the development of a new technique for exploring one of the last unobserved epochs in the history of the cosmos. We are looking for the faint radio signature of the impact of the first stars, galaxies, and black holes on the intergalactic hydrogen gas that pervades the universe. We call this period that spans from the first stars through the eventual heating and reionization of the intergalactic gas the "Cosmic Dawn." We haven't seen it yet.

It's easy to despair at the challenge of detecting that faint signal amidst contaminants orders of magnitude stronger. And it's easy to despair that the golden age is over and that all we're doing is filling in the gaps. In a sense, that's literally true. There's a blank space in our cosmic timeline and we're trying to fill it in. But in the way that really matters, I don't believe any of that. I've titled this thesis *It's Always Darkest Before the Cosmic Dawn* because, despite any occasional doubt or despair, I think what we're doing is important and stands a good chance of being something really big. We're not done yet. I'm not done yet. The search will continue.

I really believe that we're still in a golden age of cosmology. The golden age continues because the advance of our technology continues. Bigger and faster computers let us store and analyze more data. For radio astronomy, better computers leads directly to bigger and more sensitive telescopes. It's a golden age of statistics and of "big data" (whatever that means) and cosmology is fundamentally a statistical discipline. If we are, as Hubble put it, to "measure shadows" and "search among ghostly errors of measurement for landmarks that are scarcely more substantial," it sure helps to make a lot of measurements.

Big discoveries don't end golden ages—they help us to see that we're in one. Big discoveries lead to new puzzles. We need to solve puzzles to find anomalies. We need anomalies before we can have revolutions. We need revolutions for new paradigms with new puzzles to solve. To do science, one must remember that testing theories by solving puzzles and advancing revolutions by finding anomalies go hand-in-hand. It also helps to remember this. *The End of Science* was published on May 12th, 1996. David Schramm died in a tragic plane crash in December 19, 1997. Three months later, the High-z Supernova Search Team announced the discovery of dark energy.

We're not out of puzzles yet.

Chapter 1

Introduction

1.1 The Cosmic Dawn

Over its 13.8 billion year history, our universe has undergone a dramatic transformation. Just 380,000 years after the big bang, when electrons and nuclei combined for the first time and the sea of cosmic microwave background (CMB) photons decoupled from them, the universe was nearly homogenous and isotropic. Fluctuations in density and temperature were a mere part in 100,000. This exotic early universe bears almost no resemblance to today's universe, with its incredible complexity and diversity of phenomena. From the sparsest intergalactic gas to the densest cores of neutron stars, modern densities range by more than a factor of 10^{44}.

Part of that transformation was driven by the expansion of the universe, the history of which we now know very precisely. Our standard cosmological model, ΛCDM, describes a universe that is today is only 5% ordinary matter with the rest, 26% dark matter and 69% dark energy [179], made out of stuff (for lack of a better word) that we know very little about. Λ represents dark energy that acts a "cosmological constant;" it has an energy density that doesn't change as the universe expands and leads to accelerated expansion. CDM stands for cold dark matter, stuff which does not interact electromagnetically but which is massive enough and slow enough to get trapped gravitationally into halos which host modern galaxies. Along with a handful of other parameters, this cosmological model describes the expansion history of our
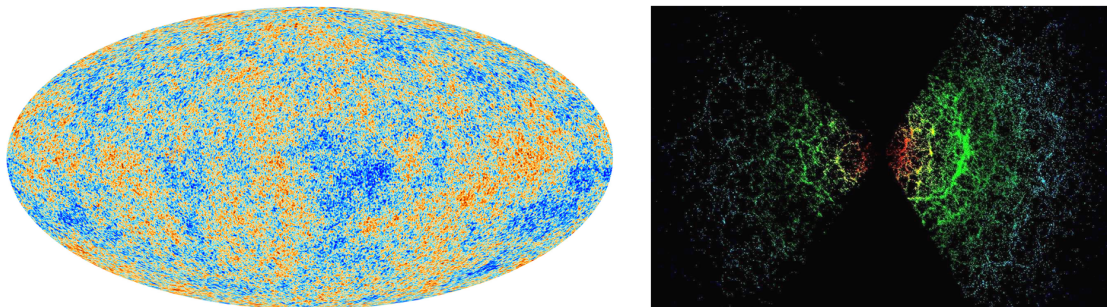
Figure 1-1: Two probes of the distribution of matter in our universe. The cosmic microwave background (left) gives us a snapshot of a nearly homogenous universe only 380,000 years after the big bang. Galaxy surveys (right) give us information about the relatively local universe, having been transformed dramatically over its 13.8 billion year evolution. *Image credits: the Planck Collaboration and the Sloan Digital Sky Survey Collaboration, respectively.*

universe very precisely and fits all available data.

If it weren't for those initial seed fluctuations in density, our universe would be far bigger and colder than it was 13.8 billion years ago, but just as boring and lifeless. Those tiny fluctuations in the density of both dark and ordinary matter evolved into stars and galaxies and planets and people. The source of those fluctuations is a great mystery, one potentially solved by invoking cosmic inflation, an early period of exponential expansion in a tiny fraction of a second.

Another daunting challenge is to explain that evolution over large time, mass, and spatial scales. Our success so far speaks volumes of the incredible progress of modern cosmology. Our understanding of the growth of structure in the universe is anchored at both ends by observation. Our record of the earliest times comes from the CMB, that thermal relic of the big bang, which we observed highly redshifted ($z \approx 1100$) by the expansion of the universe. It arrives at our telescopes today largely unperturbed by the intervening structure. In the local universe, we can probe the distribution of matter by cataloging the brightest tracers of it—namely galaxies and the supermassive blackholes they host—among other techniques (see Figure 1-1).

In between, our knowledge gets sparser, especially as we look further back in cosmic time. We're limited to observing only the brightest galaxies and active galactic nuclei (AGN) and, with some hard work, the structure along the lines of sight to
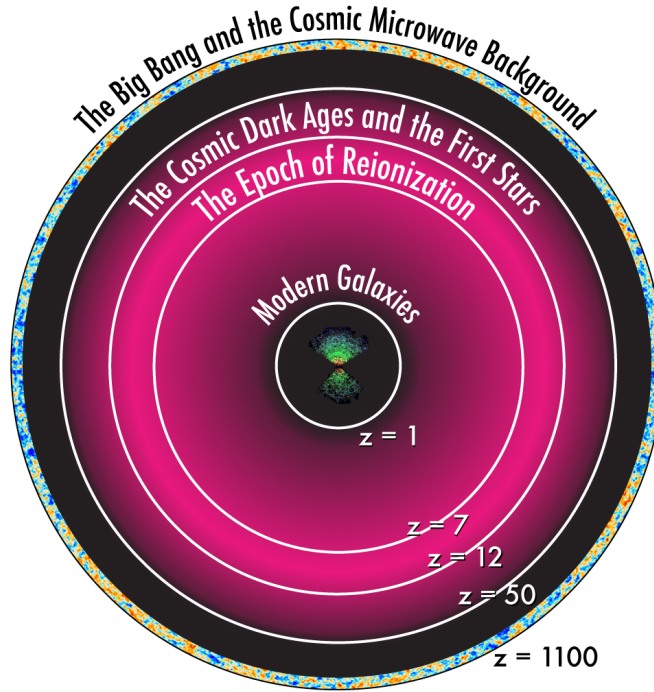
Figure 1-2: The most complete maps of the distribution of matter in the universe—CMB measurements and galaxy surveys—only probe a small fraction of the volume of the observable universe. In fact, due to its expansion, 80% of the volume of the observable universe today corresponds to regions we can see from when the universe was less than a billion years old ($z \gtrsim 6$). 21 cm cosmology, the probe that this thesis focuses on developing, may one day make the entire pink region accessible to direct observation. *Adapted from [210].*

those bright objects. As Figure 1-2 shows, an incredible fraction of the volume of the universe is unexplored, especially during the first billion years after the big bang.

As we look backwards to earlier and earlier times, we also see evidence for another dramatic transition. As dark matter halos condensed gravitationally, the ordinary matter they host cooled and collapsed to form the first generation of stars and galaxies. The formation of these first luminous objects, including the first black holes that grow by accreting matter and shining brightly in X-rays, had a dramatic effect on the rest of the universe. They heated the gas between galaxies, the intergalactic medium (IGM), and eventually reionized it. This process, starting with the first luminous objects and going through the reionization of the intergalactic gas (depicted schematically in Figure 1-3) is known as the "Cosmic Dawn."
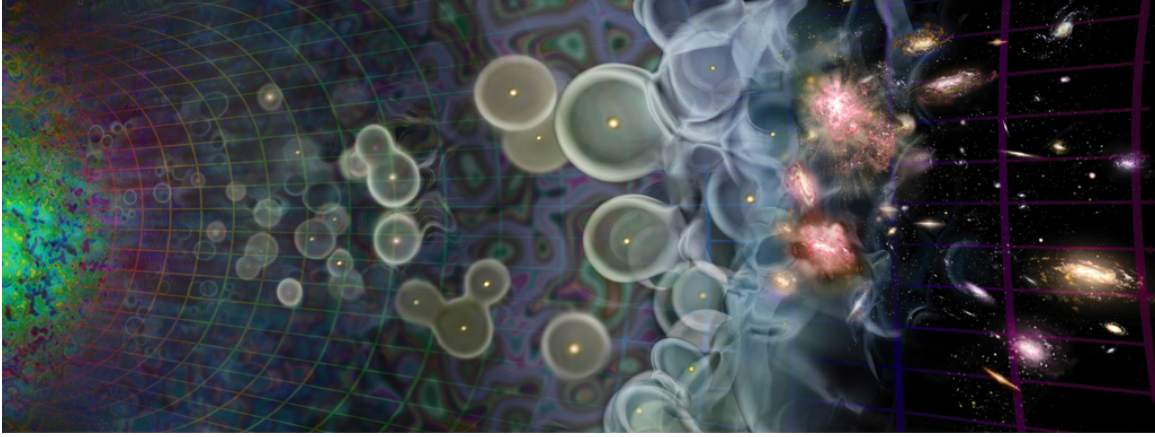
Figure 1-3: The "Cosmic Dawn" is a period in the early history of the universe between the cosmic microwave background (on the left) and modern stars and galaxies (on the right). During this time, the first stars and galaxies form and eventually heat up the intergalactic gas. They also begin to ionize the gas around them, inhomogenously filling the universe with merging bubbles of ionized hydrogen. *Image credit: Abraham Loeb and Scientific American.*

Precious little is known about exactly how this process proceeded. This thesis is devoted to advancing a new probe of the Cosmic Dawn known as "21 cm Cosmology." Before I explain the theory behind 21 cm cosmology in Section 1.2 and the ongoing observational efforts in Section 1.3, I will briefly review what we do and don't know about the Cosmic Dawn.

1.1.1   What Do We Know?

Since we can observe the universe before and after the Cosmic Dawn, it's fair to say that we know its basic story from ΛCDM. Dark matter halos collapsed, starting with the smallest overdensities and growing hierarchically. They played host to the collapse and cooling of gas into the first generation of stars and galaxies which eventually heated the IGM and reionized the universe. To try to confirm the basic story that we see play out in our simulations, we do have a few indirect observations.

First, we know that the universe finished reionization when it was about a billion years old (redshift $z \approx 6$). That information comes from observing the absorption of the first electronic transition of hydrogen, the Lyman-alpha line, in the spectra of high redshift AGN. Even a small amount of neutral hydrogen along the line of sight com-

pletely saturates the absorption signature, creating a spectral feature known as the Gunn-Peterson trough [17]. While these observations fix the endpoint of reionization and the Cosmic Dawn, they can't tell us much about the process itself.

From the cosmic microwave background, we can get a rough sense of the midpoint of reionization. The amplitude of the fluctuations (see Figure 1-1) is affected by the scattering of CMB photons off free electrons along the line of sight. The longer the universe has been ionized, the larger the damping of fluctuations in the CMB. Combining those measurements with large-scale fluctuations in the polarization of the CMB, the total level of scattering corresponds to a midpoint redshift of $z = 8.8$, assuming instantaneous reionization [179]. Of course, reionization didn't happen instantaneously across the entire universe, but the integrated constraint from the CMB is a useful starting point.

Lastly, we can make some inferences about the few galaxies we can see at very high redshift. Observations of the Hubble Ultra Deep Field, tell us about the abundance of the absolute brightest end of the distribution of galaxy luminosities up to about redshift $z = 10$ [133]. Since young, massive, and short-lived stars dominate the production of ionizing photons, the star formation rate in these galaxies should hold an important clue as to how the universe reionized. Extrapolating from a relationship that relates infrared and ultraviolet brightness in the local universe to measured star formation rates, Robertson et al. [190] find that, with reasonable modeling assumptions and an extrapolation down to very faint galaxies, those observations are consistent with the reionization inferred from the CMB. All of that is still very uncertain and model-dependent, but it confirms that the basic story is plausible.

1.1.2 What Don't We Know?

Though we have a plausible picture of reionization, the exact set of astrophysical processes that drove it are still largely unconstrained. We would like to know:

- When exactly did reionization happen and how long did it take?

- How did early stars and galaxies affect the IGM before reionization?

- When did the first stars form?

- How were they different from later generations of stars?

- What role did early black holes in X-ray binaries play in the thermal and ionization history of the IGM?

- What role did the accretion of matter onto growing supermassive black holes play?

- Which galaxies were responsible for reionization?

- How did reionization affect any subsequent formation of stars, especially in dwarf galaxies?

All of these, and many more, are important unresolved questions that we want to answer to verify our general picture of the formation of structure in the universe. We have a general outline and it is essential to see if all the details fit together.

If we find inconsistencies between our observations and our theories of the astrophysical processes that describe the formation of early stars and galaxies and their interaction with the IGM, that would be very interesting. But what makes cosmology so exciting is the potential for surprising new discoveries that would completely change our understanding of the cosmos. By exploring huge fraction of the volume of the universe only accessible by looking back to the Cosmic Dawn, we can perform very sensitive tests of our standard model of cosmology. For example:

- Measuring the statistical distribution of matter during the Cosmic Dawn via the matter power spectrum could provide extremely precise constraints on ΛCDM parameters, or reveal inconsistencies between the model and observations [135]. Of particular interest are the possible constraints—both from the power spectrum and from higher order statistics—on the simplest models of inflation, the source of those primordial density perturbations.

- Another test of ΛCDM picture and inflation would be to look for the effect of the relative velocities of dark matter and ordinary matter, which should show

up in sufficiently sensitive statistical probes of the Cosmic Dawn [226].

- The strength of primordial magnetic fields, also a prediction of inflation, could be constrained by their effect on the thermal history of the IGM [186].

- "Warm" dark matter models, where the dark matter particle was light enough to remain relativistic for much of the history of the universe, tend to wash out structure on small scales. If small fluctuations are responsible for early heating, this would affect the thermal and ionization history of the IGM [147]. Warm dark matter is currently a topic of great interest because it would help explain some potential discrepancies in the comparison of simulations observations of galaxy formation and because of a reported X-ray spectral line consistent with decaying warm dark matter [32].

- Likewise, there's also interest in recent reports of a gamma ray signature from the galactic center consistent with cold but annihilating dark matter [52]. The annihilating dark matter may also alter the thermal history of the IGM [147], meaning that observations of the Cosmic Dawn may prove a sensitive test of these theories.

To begin to answer all these questions and test these ideas with precision measurements, we need new ways to directly probe the universe during the Cosmic Dawn.

1.2   The Promise of 21 cm Cosmology

While it is exceedingly difficult to observe the first stars and galaxies to form in our universe directly, that doesn't mean that the Cosmic Dawn is unobservable. Instead of studying the brightest early objects, we can study the gas between them, the IGM. The IGM plays a fundamental role in the development of structure, since it is the source of fuel for early star-forming galaxies and it is dramatically impacted by their evolution. It is not surprising therefore that the density, temperature, and ionization evolution of the IGM across cosmic time encodes considerable information about our universe and the Cosmic Dawn.

Our best hope to directly observe the IGM during the Cosmic Dawn is to look for the radio signature of neutral hydrogen. Its ground state is very slightly split into two energy levels related to the relative spins of the proton and the electron. The incredibly precisely measured transition between the states two corresponds the emission or absorption of a photon with a frequency of $\nu_0 = 1420.40575177\,\mathrm{MHz}$, or a wavelength of about $21\,\mathrm{cm}$.

Since we understand the expansion of the universe very well, we can directly relate radio maps at different frequencies to different redshifts and thus different distances from us. Multi-frequency maps—which are comparatively easy to produce with low-frequency radio telescopes—thus represent large 3-D volumes of the universe. In this way, we can build up enormous tomographic maps, one frequency at a time. An enormous volume of the universe may be observable with these techniques (see Figure 1-2).

The scientific potential of these maps is tremendous. As I will discuss in Section 1.2.2, the huge volume of the universe accessible will enable precise tests of ΛCDM. At the same time, they will also provide the first direct observations of the astrophysical processes that drove the Cosmic Dawn.

In this section, I will review the physical processes that create the $21\,\mathrm{cm}$ signal and make it visible against the backdrop of the CMB (Section 1.2.1). Then I will review how the $21\,\mathrm{cm}$ signal is expected to vary across cosmic time and how that will translate into statistical probes of neutral hydrogen in the high-redshift IGM (Section 1.2.2).

1.2.1 The Astrophysics of Neutral Hydrogen Cosmology

The 21 cm transition has been astrophysically useful since it was first observed in 1951 by Ewen and Purcell [64]. It can be used to trace neutral gas in nearby galaxies and to measure their rotation curves. In the local universe, 21 cm emission can only be seen in galaxies where gas can cool enough to form neutral hydrogen and where the gas is dense enough that it is effectively shielded from the ionizing background. Before and during reionization, the IGM can be observed in 21 cm emission or absorption

relative to the CMB. In this section, I will explain the reasons why the 21 cm signal is visible relative the CMB and how it traces ionization, temperature, and density fluctuations in the IGM.

In radio astronomy, we typically measure the specific intensity of emission at the frequency ν, I_ν. At frequencies much lower than the peak of the CMB, we can use the Rayleigh-Jeans limit of the blackbody spectrum to represent observed intensities as brightness temperatures T_b, where

$$I_\nu \equiv \frac{2k_B T_b \nu^2}{c^2}. \tag{1.1}$$

In this limit the equation of radiative transfer through a cloud of hydrogen backlit by the CMB can be written [74] as

$$T_b(\nu) = T_S \left(1 - e^{-\tau_\nu}\right) + T_\gamma(z) e^{-\tau_\nu}. \tag{1.2}$$

Here $T_\gamma(z)$ is the temperature of the CMB at the epoch considered, τ_ν is the optical depth of the cloud due to the 21 cm transition, and T_S is the spin temperature of the gas. The spin temperature, which is the excitation temperature of the hyperfine transition, is defined in terms of the Boltzmann factor for the spin-singlet and spin-triplet hyperfine levels of the ground state of hydrogen,

$$\frac{n_{\text{triplet}}}{n_{\text{singlet}}} = 3e^{-h\nu_0/k_B T_S}. \tag{1.3}$$

The factor of 3 comes from three-fold degeneracy of the triplet state (hence the name).

The 21 cm transition is highly "forbidden" quantum mechanically, leading to a calculated lifetime for spontaneous emission of about 3×10^7 years [74], making τ_ν small and the entire IGM optically thin. It follows then that contrast in the 21 cm

signal observed today relative to the CMB, δT_b^{obs} is given by

$$
\begin{aligned}
\delta T_b^{\text{obs}} &= \frac{T_b(z)}{1+z} - T_\gamma(z=0) \\
&= \frac{(T_S - T_\gamma(z))\left(1 - e^{-\tau_{\nu_0}}\right)}{1+z} \\
&\approx \frac{T_S - T_\gamma(z)}{1+z} \tau_{\nu_0}.
\end{aligned}
\tag{1.4}
$$

I omit here the detailed calculation of the optical depth integrated over frequency to get τ_{ν_0} and, following Furlanetto et al. [74] and Pritchard and Loeb [188], simply state the final result:

$$
\delta T_b^{\text{obs}}(\hat{\mathbf{r}}, z) \approx (27\,\text{mK}) x_{\text{HI}} \left(\frac{T_S - T_\gamma(z)}{T_S}\right) (1 + \delta_b) \sqrt{\frac{1+z}{10}} \left[\frac{(1+z)H(z)}{\partial v_\parallel / \partial r_\parallel}\right].
\tag{1.5}
$$

Of course, as we observe in different directions $\hat{\mathbf{r}}$ or at different redshifts, we see different values of δT_b^{obs}. These fluctuations are sourced in three principal ways. First, ionization can drive the neutral fraction, x_{HI}, to from 1, fully neutral, to 0, fully ionized with no 21 cm signal at all. The second is due to spin temperature fluctuations relative to the CMB temperature as a function of time and position. When $T_S \gg T_\gamma$, this term saturates. However, when T_S is very cold, this can drive the signal into strong absorption relative to the CMB. Third, baryon over-densities δ_b lead to stronger signals. The last factor in Equation 1.5 comes from the Doppler broadening of the 21 cm line, which depends on the Hubble factor, $H(z)$, and gradient of the proper velocity along the line of sight, $\partial v_\parallel / \partial r_\parallel$, which includes both the Hubble expansion and the peculiar velocity of the gas cloud [74].

It is clear from Equation 1.5 that the spin temperature plays a key role in determining the observability of the 21 cm signal. If T_S is in equilibrium with T_γ, then $\delta T_b = 0$. If $T_S \ll T_\gamma$, then the 21 cm signal shows up very strongly in absorption. T_S is determined [74, 188] by the interplay of three processes:

- CMB photons at or near the 21 cm transition can be absorbed or lead to stimulated emission. This couples T_S to T_γ.

38

- Collisions between neutral hydrogen atoms and other particles may induce exchanges of angular momentum, causing a spin flip. This effect is dominated by hydrogen-hydrogen collisions, hydrogen-electron collisions, and hydrogen-proton collisions, all of which couple T_S to the kinetic gas temperature, T_K.

- Absorption and remission of Lyman-alpha photons allows an indirect path to changing the hyperfine state of hydrogen, since transitions from the 1S state of hydrogen to some of the 2P states and back allow a net spin flip. This couples T_S to T_α, the color temperature of the Lyman-alpha transition, defined analogously to Equation 1.3. This pathway for hyperfine transitions is known as the *Wouthuysen-Field effect* [238, 66].

In equilibrium, the spin temperature is given by

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_c T_K^{-1} + x_\alpha T_\alpha^{-1}}{1 + x_c + x_\alpha}, \tag{1.6}$$

where x_c and x_α, the collisional and Lyman-alpha coupling coefficients depend on the subtle atomic processes that govern these effects, which themselves have complicated temperature and density dependences [74, 188].

1.2.2 The 21 cm Signal Across Cosmic Time

The physical processes that drive ionization, spin temperature, and density changes that create δT_b^{obs} in Equation 1.5 are both inhomogenous and time-dependent. Across cosmic time, the 21 cm signal and its underlying statistics are expected to change dramatically, though the precise evolution depends on the poorly understand processes that drove the Cosmic Dawn.

In the top panel of Figure 1-4, I show one possible history of δT_b^{obs}, reproduced from Pritchard and Loeb [188]. We can see readily that the evolution of the brightness temperature is complicated and markedly different during different epochs. Fully extracting cosmological and astrophysical information from this process requires large, detailed maps across many redshifts.
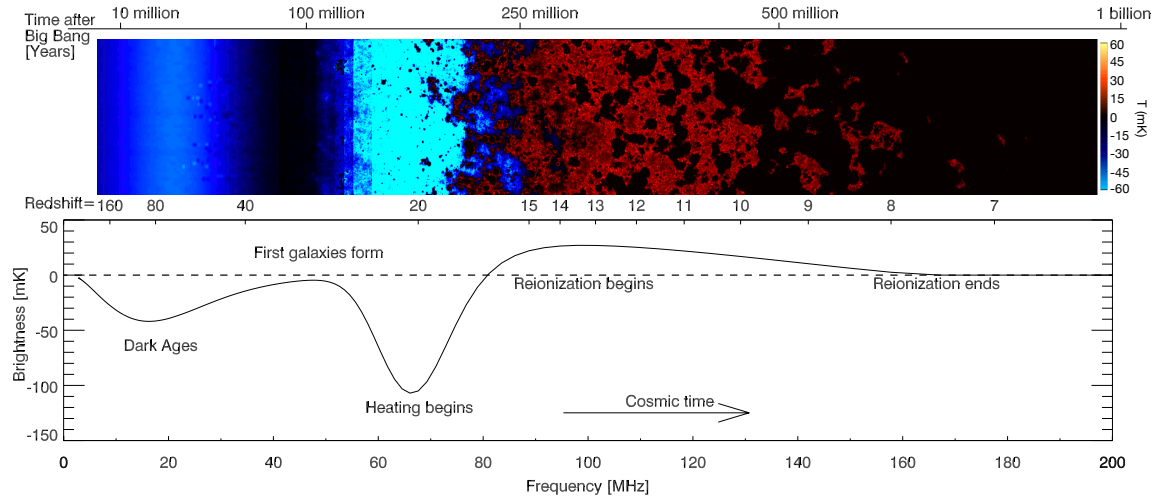
Figure 1-4: The 21 cm brightness temperature evolves in complex ways over the course of the dark ages and the Cosmic Dawn. Different physical processes at different times cause it to appear either in absorption or in emission in constrast to the CMB, sometimes globally and sometimes inhomogenously. *Top panel:* one slice through a simulation shows the evolution of the brightness temperature of the signal and the patchy heating and ionization caused by the first generation stars and galaxies. *Bottom panel:* the sky-averaged global 21 cm signal, which largely traces the evolution of the spin temperature and neutral fraction of hydrogen before and during the Cosmic Dawn. *Reproduced from [188].*

Such maps are very difficult to produce and interpret, as I will discuss in Section 1.3, so it is useful to consider reduced data products that take advantage of the approximate statistical isotropy of the signal. The simplest statistical description of the evolution of $\delta T_b^{\mathrm{obs}}$ is the sky-averaged global signal. The global signal, plotted in the bottom panel of Figure 1-4, is expected to go through peaks and troughs as the spin temperature and ionization fraction evolve before and during the Cosmic Dawn.

Another useful way to statistically probe the 21 cm signal would be to look for correlations on particular length scales. During reionization, for example, we expect correlations on the characteristic length scale associated with growing ionized bubbles around early galaxies. This quantity is most conveniently represented in Fourier space as the power spectrum, $P(\mathbf{k})$, where

$$\langle \widetilde{\delta T_b^*}(\mathbf{k})\widetilde{\delta T_b}(\mathbf{k}')\rangle \equiv (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}')P(\mathbf{k}), \tag{1.7}$$

where angle brackets denotes an ensemble average, $\widetilde{\delta T_b}(\mathbf{k})$ is the Fourier transform of $\delta T_b(\mathbf{r})$, and $\delta(\mathbf{k} - \mathbf{k}')$ is the Dirac delta function. If the 21 cm signal is statistically isotropic—which should be a good approximation—then $P(\mathbf{k})$ reduces $P(k)$. Often the power spectrum is reported as a "dimensionless" power spectrum[1] $\Delta_{21}^2(k)$ where

$$\Delta_{21}^2(k) \equiv \frac{k^3}{2\pi^2}P(k). \tag{1.8}$$

Because the 21 cm signal is not a Gaussian random field, the power spectrum does not contain all of the cosmological information in the maps themselves. But by measuring just a few values of the power spectrum as a function of k and z, we can extract much of the available information while significantly reducing the noise on our final measurements. Most of this thesis is concerned with the estimation of the 21 cm power spectrum, both in theory and in practice, and how it can be used to constrain the physics behind the Cosmic Dawn.

In the remainder of this section, I will briefly summarize the theorized stages in

---

[1]For the brightness temperature power spectra we measure in 21 cm cosmology, it actually has units of temperature squared.

the evolution of the 21 cm signal and their observable statistical properties. Further information on these processes can be found in Pritchard and Loeb [188].

1.2.2.1 High Redshifts

The 21 cm signal first becomes distinguishable from the CMB around $z \approx 200$. Before that redshift, residual free electrons couple the gas kinetic temperature to the CMB temperature, setting both T_K and T_S to T_γ. Around $z = 200$, this process is no longer effective and the gas begins to cool adiabatically. Therefore, while the temperature of the CMB goes as $T_\gamma \propto (1+z)$, the gas cools like $T_K \propto (1+z)^2$. As long as collisional coupling is effective, which it is thought to be until $z \approx 40$, this sets $T_S < T_\gamma$ and makes the signal appear in absorption. This process accounts for the first dip in the global signal in Figure 1-4. Since T_S is fairly uniform during this period and $x_{\mathrm{HI}} \approx 1$, spatial fluctuations in the 21 cm signal are sourced by density fluctuations alone. Being able to observe these fluctuations would provide a spectacularly clean probe of the matter power spectrum and a precise test of ΛCDM, though observations at this redshift are well beyond the limits of current technology.

The second dip in the global signal is caused by the combination of two processes. As the first stars in the universe form, they produce enough Lyman-alpha photons to couple T_S to T_α via the Wouthuysen-Field effect. Since the universe is mostly neutral and the optical depth to Lyman-alpha in the IGM is very large, T_α is driven toward T_K, which is less than T_γ. That causes the 21 cm signal to be visible again in absorption. Fluctuations in the 21 cm field are caused by variations in Lyman-alpha field corresponding to the first dark matter halos to collapse and form stars. Eventually, heating of the IGM by X-ray sources, like the first X-ray binaries and micro-quasars, drives T_K above T_γ and the 21 cm signal into emission. Since this process happens inhomogeneously, it is expect that that the signal will be visible in emission in parts of the sky and absorption in other parts of the sky simultaneously, potentially leading to observable effects in the 21 cm power spectrum (see Chapter 7). This "Epoch of X-ray Heating" drives $T_K \gg T_\gamma$, saturating the T_S term in Equation 1.5.
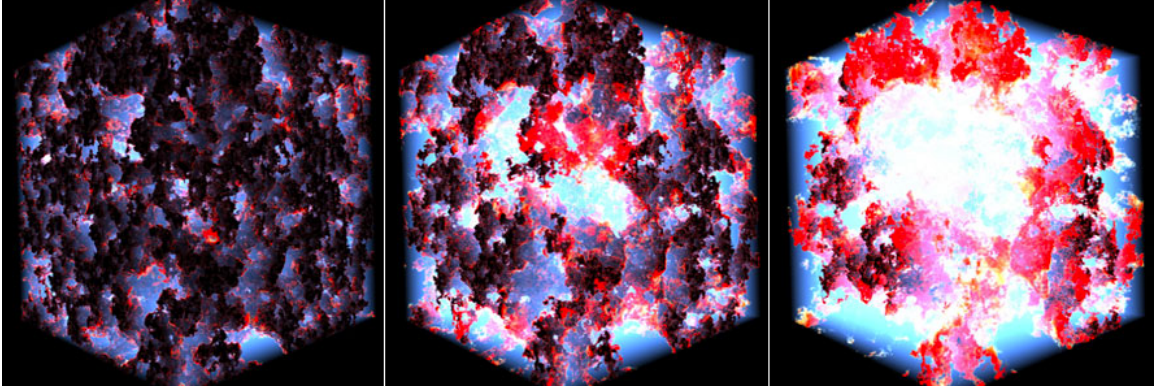
Figure 1-5: A simulation of reionization in a cube approximately 300 Mpc on a side. As reionization proceeds small bubbles of ionized hydrogen (bright areas) grow to eventually dominate the neutral gas (dark areas) and completely ionize the IGM. *Image credit: Marcelo Alvarez, Ralf Kaehler, and Tom Abel.*

1.2.2.2   The Epoch of Reionization

Around that time, reionization of the intergalactic medium by ultraviolet photons from young, high-mass stars is expected to begin, leading to growing bubbles of ionized gas around early galaxies. As the simulation in Figure 1-5 shows, ionized bubbles eventually grow and coalesce. This reduces the fraction of neutral hydrogen and thus the strength of the 21 cm global signal.

If the spin temperature at reionization is far larger than the temperature of the CMB, then variations in $\delta T_b^{\rm obs}$ are created by density and ionization fluctuations, the later of which evolved dramatically over the course of the EoR. At the beginning of reionization, density fluctuations determine the 21 cm power spectrum, leading to higher power at high k in $\Delta_{21}(k)$. As the ionized bubbles grow, they erase very small scale (high k) fluctuations but create correlations on large scales (low k). This is reflected in the expected evolution of the 21 cm power spectrum in Figure 1-6. As reionization proceeds, the overall amplitude of the power spectrum decreases because it is proportional to $x_{\rm HI}^2$. But we also see the formation of the "knee" in the power spectrum that moves to lower k as the characteristic bubble size increases.

Simulations of the 21 cm power spectrum [141] have found that it depends more strongly on $x_{\rm HI}$ than on the redshift of reionization. It follows them that the power spectrum will be a sensitive probe of the ionization history of the universe, which is
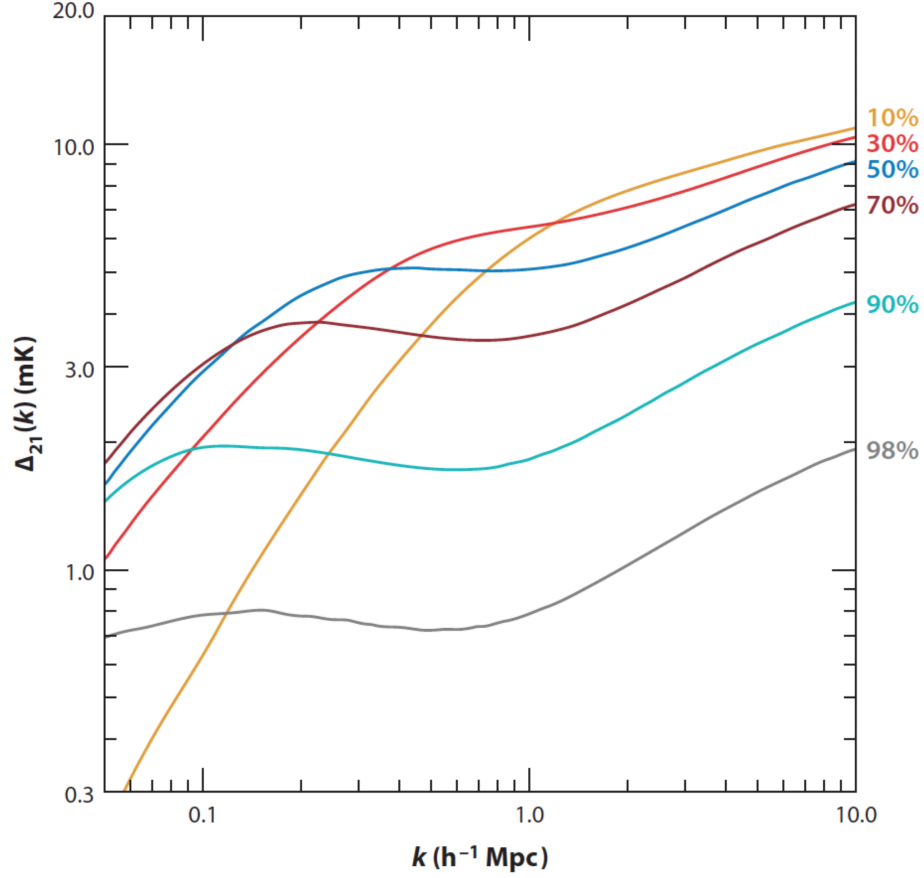
Figure 1-6: The evolution of the 21 cm power spectrum with ionized fraction is expected to reveal a tremendous amount of information about the processes that drove reionization and the physics of the first stars and galaxies. As the universe goes from mostly neutral (yellow) to mostly ionized (gray), the overall amplitude of the power spectrum is expected to decrease, since x_{HI} normalizes δT_b. However, the growth of ionized bubbles creates correlations on the characteristic size scale of those bubbles, increasing low k power during the early stages of reionization. *Reproduced from [154].*

still largely unknown.

The exact shape of the power spectrum and its evolution with x_{HI} or z depends sensitively on the astrophysics of reionization. In Chapter 8 my collaborators and I examine the qualitative differences between power spectra when varying parameters of a relatively simplistic reionization model. Because the 21 cm power spectrum varies so dramatically over cosmic time as a function of k, it can be used to sensitively probe the physics that drove it. Specifically, we found that a next-generation telescope could constrain these parameters at roughly the 5% level using the power spectrum. Though we know that reionization was over by $z \approx 6$, we don't know exactly when it began or how long it took. Thus, observations aimed at this signal usually observe at $13 \lesssim z \lesssim 6$, corresponding to a frequency between 100 and 200 MHz.

Of course, part of the promise of 21 cm cosmology is that it makes an enormous volume of the universe accessible to observation, providing an exquisite test of ΛCDM and possible extensions to it. If $P(\mathbf{k})$ is decomposed into powers of μ where $\mu \equiv \mathbf{k} \cdot \hat{\mathbf{r}}$, it can be shown from linear perturbation theory that the μ^4 term depends only on density fluctuations [12, 13]. With a large enough telescope optimized for 21 cm cosmology, Mao et al. [135] showed that 21 cm power spectra measured over a fairly large range of redshifts can reduce the errors on cosmological parameters like Ω_Λ, Ω_m, Ω_b, n_s, Ω_k, and $\sum m_\nu$ by an order of magnitude or more compared to what's possible with current CMB observations. While these measurements are still rather futuristic, they serve as a shining example of what's possible with 21 cm tomography.

1.2.2.3 Low Redshifts

Though hydrogen in the IGM was completely ionized by $z \approx 6$, galactic halos can still host residual neutral hydrogen where densities are high enough that recombination rates exceed ionization rates, shielding the neutral gas. While it will be very difficult to observe individual galaxies, low resolution images that average together emission from many galaxies may enable a measurement of the underlying matter power spectrum. However, this requires modeling the bias factor that relates dark matter halos to the amount of neutral hydrogen that they host, which may vary as a function of galaxy

mass, size, and age in non-trivial ways.

More promising is the ongoing effort to measure baryon acoustic oscillations in the power spectrum [241, 38, 8]. Since the baryon acoustic scale serves as a standard ruler, measuring it in the 21 cm power spectrum as a function of z can constrain the expansion history of the universe and thus the dark energy equation of state. Since the acoustic scale at 150 Mpc is much larger than individual galaxies, the difficulty of measuring the signal from individual galaxies is less important than ease of building sensitive telescopes with wide fields of view and precise redshift information. Unlike with optical and infrared surveys that have measured the baryon acoustic signal, 21 cm "intensity mapping" experiments get redshift information basically for free, potentially making cosmic-variance-limited measurements relatively inexpensive.

1.3   Observational Challenges of 21 cm Cosmology

Though a detection and characterization of the 21 cm signal from the epoch of reionization would be an invaluable tool for understanding our Cosmic Dawn, actually making the measurement has proven extremely difficult. In fact, Parts I and II of this thesis are devoted to exploring and overcoming both the theoretical and real-world challenges of making a detection. In this section I will review the basics of interferometry (Section 1.3.1), how we plan to separate out astrophysical foregrounds that are many orders of magnitude stronger than the cosmological signal (Section 1.3.2), and the current (Section 1.3.3) and next generation (Section 1.3.4) efforts to detect the 21 cm signal.

1.3.1   Low Frequency Radio Interferometers

Unlike traditional telescopes that measure energy deposited in a focal plane, radio telescopes measure incident electric fields from the sky directly. If we make the generally very accurate approximation that radio emission from different sources on the sky is incoherent, then it follows that the correlation of measurements from different antennas can tell us about what's on the sky. We call this time-averaged correlation
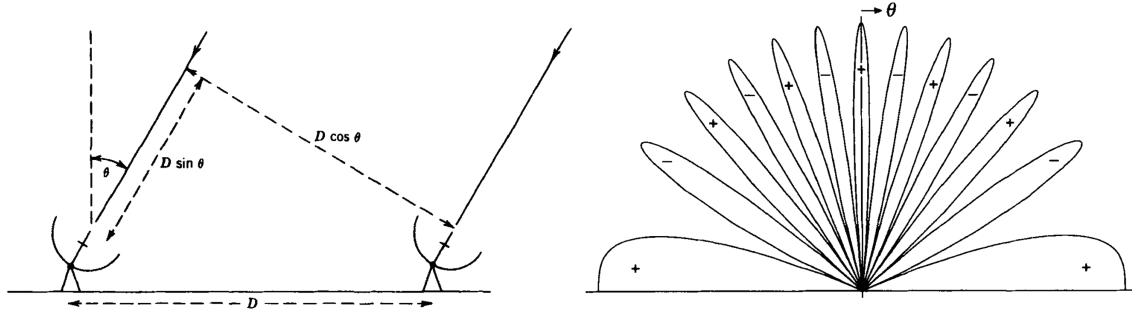
Figure 1-7: By correlating the signals from two spatially seprated antennas, radio interferometers are sensitive to the delay between when light from a source arrives at one antenna and when it arrives at the other (see the lefthand panel). These correlations, called "visibilities," are weighted measurements of Fourier modes on the sky (see the righthand panel). With many such measurements with a variety of antenna separations and relative orientations, images of the sky can be reconstructed with high sensitivity. *Reproduced from [216].*

between signals measured at antenna i and antenna j the "visibility," V_{ij}. It's given by[2]

$$V_{ij}(\nu) = \int B_{ij}(\hat{\mathbf{r}}, \nu) I(\hat{\mathbf{r}}, \nu) \exp\left[-2\pi i \frac{\nu}{c} \mathbf{b}_{ij} \cdot \hat{\mathbf{r}}\right] d\Omega. \qquad (1.9)$$

This equation can be interpreted as saying that a pair of antennas displaced by vector $\mathbf{b}_{ij}$ are sensitive to the sky, $I(\hat{\mathbf{r}}, \nu)$, weighted by the product of the sensitivities of the antennas, $B_{ij}(\hat{\mathbf{r}}, \nu)$, also known as the "primary beam." However, the correlation between the signals from two antennas is only observed with an extra time-delay corresponding to the separation between antennas along the line of sight to a source (see the lefthand panel of Figure 1-7). This extra time delay introduces the phase factor in Equation 1.9.

As a result of that phase factor, visibilities really measure Fourier modes of the beam-weighted sky. Parts of the sky interfere constructively, other parts destructively, as the righthand panel of Figure 1-7 illustrates. A pair of antennas can be very sensitive to changes in position perpendicular to their orientation, since that can rapidly change the phase factor. If the antennas are nearby, or if position changes are perpendicular to their separation, the phase changes slowly.

---

[2]For this discussion, I ignore the complications that arise when measuring a polarized signal. A more complete treatment can be found in Thompson et al. [216].

This can be generalized. With N antennas, we can measure $N(N-1)/2$ different visibilities. As the Earth rotates, $\mathbf{b}_{ij} \cdot \hat{\mathbf{r}}$ changes, allowing for the measurement of new Fourier modes. With enough independently measured Fourier components of the sky, an image can be reconstructed via "aperture synthesis." These measure the sky convolved with a point spread function (PSF) or "synthesized beam" related to the observed antenna separations or "baselines."

Typically, astronomers build interferometric telescopes because they are interested in making measurements with very high angular resolution. Roughly speaking, the angular resolution of an interferometer is set by $\lambda/b_{\max}$, the ratio of the wavelength observed to the longest baseline. For 21 cm cosmology, our aim is not angular resolution—we get most of our sensitivity to small spatial scales from spectral resolution—but high sensitivity and large fields of view. The cost of large, single-dish radio telescopes usually scales with the collecting area as $A^{1.35}$ [210]. The physical hardware cost of building a radio interferometers scales only linearly with the collecting area, since more antennas yield more sensitivity. The computing cost of performing the correlation between antennas to calculate visibilities usually scales as N^2 and for large enough N, it can be a limiting factor. This is not true for all interferometers, as I will explain in Chapter 6.

High sensitivity is extremely important for 21 cm cosmology precisely because the 21 cm signal is so weak compared to the astrophysical foregrounds, as I will discuss in Section 1.3.2. Since most of the signal measured by a radio antenna comes from incoherent sky signals, the noise in a visibility is set by T_{sky}, which is roughly the average sky brightness temperature. T_{sky} sets the system temperature, T_{sys} because it is usually hundreds of Kelvin at EoR frequencies and thus dominates over the electronic noise in the receiver. The relationship between noise in a visibility and noise in the power spectrum is discussed in Chapters 2 and 3. Suffice it to say that first generation instruments (which I will discuss in greater detail in Section 1.3.3) likely need a thousand or more hours of observation to make a confident detection of the EoR signal [151, 25, 117, 87, 170, 15, 184]. Thus, the need for large collecting areas, combined with the relative inexpensiveness of individual antenna elements designed

to operate at low frequencies, has driven the field toward interferometers.

1.3.2 The Problem of Foregrounds

Astrophysical foregrounds remain the most daunting challenge for 21 cm cosmology. The brightness temperature we measure on the sky inevitably contains both the 21 cm signal from the Cosmic Dawn and relatively nearby, radio-bright objects that fill the entire sky at the angular resolution of our instruments. Our hope of separating the astrophysical foregrounds relies on their spectral smoothness. Measurements of CMB anisotropies faced a similar problem; they also contained smooth spectrum foregrounds much brighter than the signal they sought. In the case of the CMB, measurements at different frequencies have the same thermal blackbody signal and the same foreground contaminants. The strategy for the CMB was to look at different frequencies to differentiate the two based on their frequency dependence. In the case of 21 cm cosmology, each frequency probes an entirely new cosmological signal. That's the whole point. The ability for tomography to explore a vast volume is also the reason why the problem of foregrounds is so difficult. We need new approaches which take their cues from previous work on the CMB but must be adapted to the thornier problem at hand. In this section, I will explain what the foregrounds are, how they appear in our measurements, and what we can do about them.

1.3.2.1 What Are the Foregrounds?

At the frequencies of interest, the dominant foregrounds are synchrotron emission from our Galaxy and other radio galaxies. Synchrotron emission from our Galaxy— the result of ultrarelativistic charged particles bending in the Galaxy's magnetic fields—has some spatial structure, but is highly spatially correlated, as I show in Figure 1-8. Free-free emission also contributes, albeit at a much lower level [74]. Both sources produce very spectrally smooth foregrounds because of the physical mechanism behind synchrotron and free-free emission.

Additionally, bright radio galaxies, which are usually unresolved by our instru-
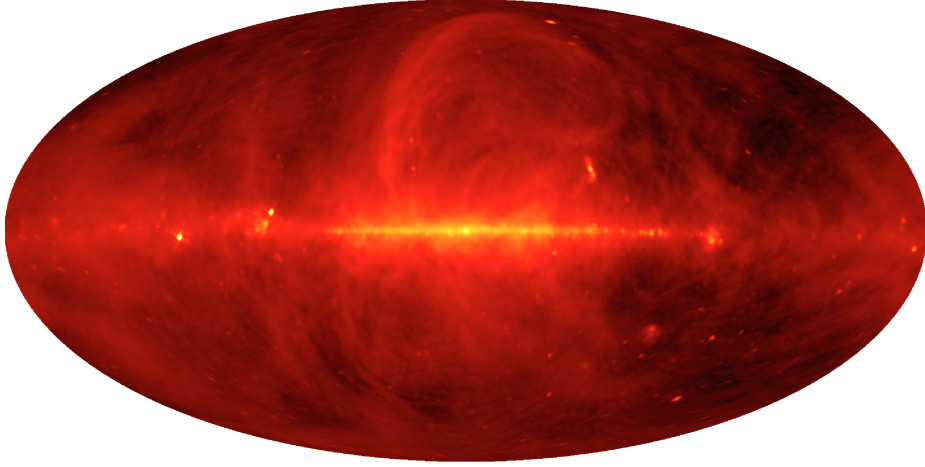
Figure 1-8: A map of diffuse radio emission, mostly from our Galaxy, at 150 MHz. At EoR frequencies, the galaxy is hundreds of Kelvin, compared to the cosmological 21 cm signal which is likely less than 10 mK. *Produced using the results of [54].*

ments, contribute considerable flux. They are generally sourced by the interaction of jets from active galactic nuclei with the surrounding IGM. They too are synchrotron dominated and are therefore spectrally smooth. Many sources contaminate every pixel of our maps and create a confusion-limited sea of unresolved point sources.

Because the dominant foregrounds are driven by processes that create inherently spectrally smooth emission, they can be well-characterized using maps at just a few frequencies. When we make maps at hundreds of frequencies, as we often do in 21 cm tomography, we can expect only a small fraction of the total information about the cosmological signal to be completely lost due to foreground uncertainty [121].

There are also foregrounds that are not so spectrally smooth. Man-made radio frequency interference (RFI) can be even brighter than the astrophysical foregrounds, but can usually be isolated in time and frequency and mitigated by building arrays at remote sites. Polarized foregrounds, if they leak into maps of unpolarized emission, may also acquire spectral structure due to Faraday rotation. So far, this effect appears to be small [150].

1.3.2.2 How Foregrounds Interact with an Interferometer

An interferometer is an inherently chromatic instrument. The phase term in Equation 1.9 depends frequency, so we should expect that the PSF or synthesized beam should also depend on frequency in non-trivial ways. The primary beam is also frequency dependent, so PSFs can vary spatially as well. Taking this into account properly is the subject of Chapter 3.

The spatial and spectra dependence of PSFs complicates the simple story of how foregrounds can be separated from the 21 cm signal. While intrinsic foregrounds are very spectrally smooth, observed foregrounds can have complex spectral structure. With a sufficiently precise understanding of the operation of the instrument—including exquisite calibration—the complex spectral structure can be modeled with just a few foreground parameters per line of sight. But actually understanding our instrumental calibration and primary beams to the roughly 0.01% level necessary is very difficult.

So, when we make 3-D maps we expect foreground contamination at every frequency, which is a proxy for distance. The signal we'd ultimately like to measure depends only on $|\mathbf{k}| \equiv k$. However, to separate foregrounds, which behave differently along the line of sight than perpendicular to it, we form power spectra in cylindrically-averaged 2-D Fourier space, parametrized by $k_\parallel$ and $k_\perp$. Were it not for the chromaticity of the instrument, we would expect foregrounds to only contaminate the lowest $k_\parallel$ modes. But, as we can see in the 2-D power spectrum plotted in the lefthand panel of Figure 1-9, the brightest, most foreground-dominated region depends both on $k_\parallel$ and $k_\perp$.

Thankfully, the smallest scale of spectral structure the instrument can impart on a given baseline corresponds to the geometric delay associated with sources at the horizon [172]. There the phase term in Equation 1.9 is maximized. Baseline length determines angular resolution and thus spatial resolution. Therefore longer baselines probe higher $k_\perp$ modes of the 21 cm power spectrum. Likewise, since delay is a Fourier dual to frequency which is a substitute for distance in 21 cm tomography,
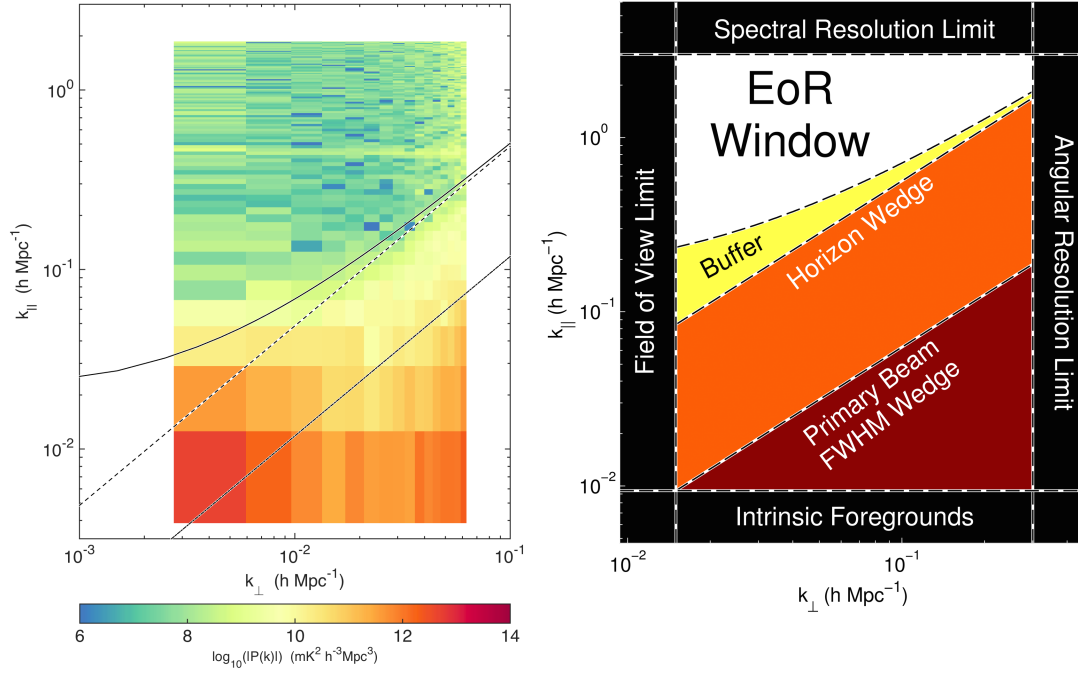
Figure 1-9: When bright but spectrally smooth astrophysical foregrounds are observed with an interferometer, an inherently chromatic insrument, they take on spectral structure. In 2-D Fourier space, where $k_\parallel$ measures Fourier modes along the line of sight and $k_\perp$ measures Fourier modes perpendicular to the line of sight, the foregrounds show up as a "wedge." That's because finer spatial scales—probed by longer baselines—have more spectral structure. The safest way to detect the 21 cm power spectrum is to work outside the wedge, in the "EoR window" (righthand panel). The EoR window has thus far proven relatively foreground free (see the lefthand panel), though working only in the window comes at the cost of sensitivity, as I will discuss further in Chapter 8. *Figures reproduced from Chapters 3 and 5, respectively.*

longer delays correspond to higher $k_\parallel$ modes. This explains the structure we see in Figure 1-9 called "the wedge," which has also been seen in simulations [50, 172, 230, 156, 89, 225, 218, 125, 126] and in observations [182, 59, 219].

Outside the wedge lies the so-called "EoR window" which should be free of foreground contamination. Exactly where the wedge-window divide occurs depends on the instrument and the foregrounds. While most interferometers are designed to have little sensitivity near the horizon, a large fraction of the total solid angle of the observable celestial sphere is near the horizon [219]. While most observed foreground emission may fall within the main lobe of the primary beam, enough foregrounds to swamp the cosmological signal may still be present in the sidelobes. If the foregrounds have some spectral structure, they are expected to leak into a buffer just beyond the wedge [182, 125], as I show in the schematic illustration of the EoR window in the righthand panel of Figure 1-9.

1.3.2.3   Two Strategies for Foreground Removal

The current leading strategy for detecting the 21 cm EoR signal relies on avoiding foregrounds by working only within the window. The current best limits in Ali et al. [6] and the strategy my collaborators and I employed in Chapters 4 and 5 used only data from inside the the window. As I will discuss in Sections 1.3.3 and 1.3.4, some telescopes are being designed to take advantage of this strategy and eschew imaging fidelity and angular resolution in favor of many short, redundant baselines that probe low $k_\perp$ modes less contaminated by the wedge.

The downside to foreground avoidance is that it sacrifices sensitivity. As my coauthors and I found in Chapter 8, giving up on Fourier modes near the edge of the wedge results in a roughly 70% drop in sensitivity even for a highly compact array. Using the yellow and perhaps even the orange modes in the righthand panel of Figure 1-9 can mean the difference between an upper limit on the 21 cm power spectrum with current generation interferometers and a solid detection.

To work in those regions, we must find a way to subtract foregrounds from our data. Foreground subtraction is very difficult and has been the subject of many papers

over the last several years (e.g. [155, 28, 122, 120, 39, 58, 40]). We need to subtract foregrounds orders of magnitude stronger than the cosmological signal which have been convolved with an instrument whose effect is only imperfectly understood. We need precise models of both foregrounds and our instrument. And most importantly, we must take our own uncertainty about these models into account. If we do not, we risk mistakenly claiming a detection. Much of this thesis (Chapters 2, 3, 4, and 5) is concerned with precisely this question: what do we need to know to subtract foregrounds and how do we translate our uncertainty about their subtraction into errors on our power spectrum measurements? The goal is to claw back as much of the EoR window as we are justified in doing, and no more.

Even if we are simply seeking to avoid foregrounds by excising the wedge region, the techniques my collaborators and I have developed are important because they can minimize the leakage of foreground power into the EoR window (see Chapters 4 and 5). Regardless of whether or not we work within the wedge, we need to know the errors on our measurements, the correlations between those errors, and the relationship of our measurements to the true cosmological $P(k)$.

Whether or not we will ever understand the foregrounds and our instruments well enough to work within the wedge is an open question. Perhaps the most important message of this thesis is that we should try to achieve the marked increase in sensitivity possible with foreground subtraction and that, even if we fail, as long as we understand our uncertainties, we'll make the best measurements that we can.

1.3.3   First Generation Interferometers and Results

The quest to detect the 21 cm signal from the epoch of reionization is well underway and a number of telescopes have set limits on the power spectrum. In this section, I'll discuss several of them, review their progress thus far, and compare their strategies for detecting the 21 cm signal from the EoR.
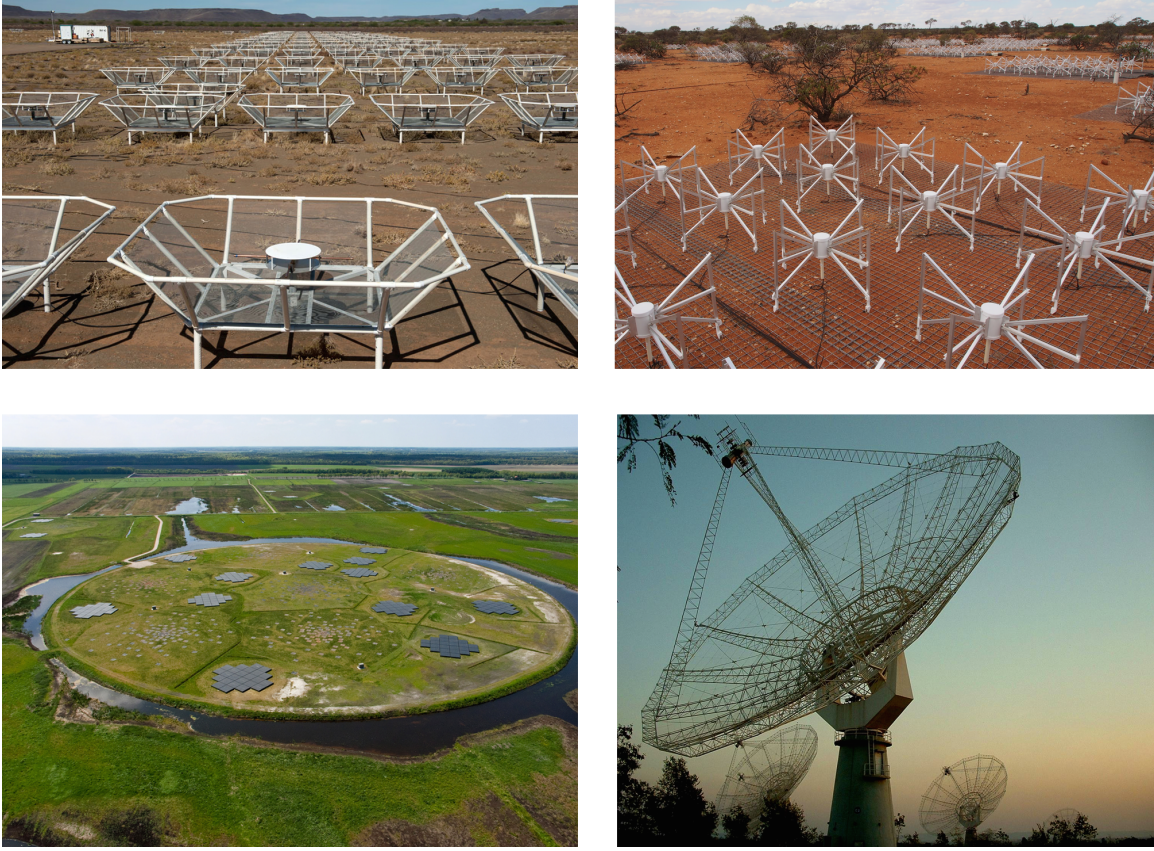
Figure 1-10: First generation interferometers trying to detect the 21 cm signal from the EoR. *Top left:* the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER). *Top right:* the Murchison Widefield Array (MWA). *Bottom left:* the Low Frequency Array (LOFAR). *Bottom right:* the Giant Metrewave Radio Telescope (GMRT). *Image credits: SKA South Africa, the SKA Collaboration, LOFAR/ASTRON, and Tzu-Ching Chang, respectively.*

1.3.3.1 Giant Metrewave Radio Telescope

The Giant Metrewave Radio Telescope (GMRT) is the oldest of the 21 cm observatories and consists of 30 steerable dishes, each 45 m in diameter (see the lower right panel of Figure 1-10). At the frequencies of interest, this yields a field of view roughly 3° across. It is a multi-purpose observatory located 80 km North of Pune, India.

The first upper limit on the 21 cm power spectrum was set with the GMRT [166], though it was later revised when it was discovered that the analysis technique for removing foregrounds also removed 21 cm signal [167]. The current best limit from GMRT is $\Delta^2(k) < 6.2 \times 10^4 \, \mathrm{mK}^2$ at $z = 8.6$ and $k = 0.50 \, \mathrm{h} \, \mathrm{Mpc}^{-1}$.

1.3.3.2 The Murchison Widefield Array

The Murchison Widefield Array (MWA), like the GMRT, is a multi-purpose observatory. However, is more focused on 21 cm cosmology than previous telescopes. It consists of 128 tiles, each made of 16 dual-polarization dipole antennas (see the upper right panel of Figure 1-10). The signals from the dipoles are added with an appropriate set of delays by an analog beamformer to focus the sensitivity of the array on particular parts of the sky. This allows the MWA to form a discrete set of primary beams on the sky, each with a full-width at half-maximum of roughly 30°. For EoR observations, this allows observers to adapt a "drift and shift" strategy, where the primary beam changes roughly once every half hour. The MWA is located in the Murchison Radio-astronomy Observatory in a remote part of Western Australia, 600 km north of Perth.

Chapter 4 of this thesis contains an analysis of 32-tile MWA prototype data. Chapter 5 updates that analysis with new foreground residual covariance modeling and applies it to 128-tile data, yielding a best (though as-yet-unpublished) upper limit of $\Delta^2(k) < 3.7 \times 10^4 \, \mathrm{mK}^2$ at $z = 6.8$ and $k = 0.18 \, \mathrm{h} \, \mathrm{Mpc}^{-1}$. Both chapters contain significantly more detail about the design and operation of the instrument. The MWA has over 1000 hours of total observation already on disk (split across two fields and two frequency bands) and analysis of deeper observations is ongoing.

1.3.3.3 The Precision Array for Probing the Epoch of Reionization

The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER) differs from competing telescopes in that it is a focused experiment designed exclusively for EoR observations. It is located in the Karoo Radio Astronomy Reserve in the Karoo Desert of South Africa. Its 128 dipoles sit atop relatively small frustrum-shaped ground screens arranged in a highly redundant configuration (see the top left panel of Figure 1-10). The redundant configuration simplifies calibration (see Chapter 6) and focuses the maximum sensitivity on a small number of baselines [172]. However, by foregoing imaging fidelity, it makes foreground avoidance the only

feasible strategy.

Despite having the least collecting area, its focused design and observing strategy have helped PAPER produce the world's best upper limits on the 21 cm power spectrum. Using only the previous 64 element configuration, PAPER set an upper limit of $\Delta^2(k) < 500\,\mathrm{mK}^2$ at $z = 8.4$ between $k = 0.15$ and $k = 0.50\,\mathrm{h\,Mpc}^{-1}$ [6]. This limit allowed the PAPER team to determine that the IGM was heated above roughly $7\,\mathrm{K}$ at $z = 8.4$, otherwise T_S would be so far below T_γ that the 21 cm signal would show up brightly in absorption [185]. Under a wide range of assumptions, achieving that level of heating requires inferring a population of high-redshift galaxies dimmer than those currently directly observed. This result is not surprising, but it one of the first constraints on the Cosmic Dawn from 21 cm observations.

1.3.3.4 The Low Frequency Array

The Low Frequency Array (LOFAR) is actually two interferometers, the High Band Array, which observes at EoR frequencies, and the Low Band Array, which was designed for other science. The High Band Array bears many similarities to the MWA in that each element of the interferometer is a analog phased array of 16 dipoles. In the LOFAR core, which is located near Exloo in the Netherlands, 24 such tiles are arranged into each of 40 "fields" (22 of which are visible in the lower left panel of Figure 1-10). Though LOFAR has a much larger collecting area than the MWA, it cannot correlate every tile with every other tile and instead generally forms beams digitally on a per-field basis, each about 10° across. Because beams are formed digitally, multiple simultaneous beams can be formed within the tile beam, though this process is limited by the tradeoff between simultaneous bandwidth and the the computing power required for correlation of what amounts to multiple interferometers simultaneously. Correlation is generally more costly for LOFAR than for PAPER or the MWA because the high level of RFI in the Netherlands necessitates very fine frequency resolution.

Thus far, LOFAR has not published any upper limits on the 21 cm power spectrum, though they have published some initial calibration, mapmaking, and source-finding

results [244]. By utilizing far-flung LOFAR stations across Northern Europe, LOFAR can achieve far higher angular resolution than other telescopes. They are attempting to use the angular resolution to subtract individual sources down past the level of the EoR signal using a number of different subtraction techniques [39, 40]; their baselines are mostly so long that foreground avoidance is too costly. The LOFAR team is also trying to measure the "variance statistic," which is effectively a power spectrum averaged over all k bins, in order to probe the redshift evolution of the cosmological signal with maximum sensitivity [174]. Interpreting that result will be more difficult than interpreting a power spectrum and it's not clear whether a measurement of the variance statistic will prove a convincing detection of the EoR.

1.3.3.5 MITEoR

Though it was not designed to have enough sensitivity to detect the EoR, the MIT EoR experiment (or "MITEoR" for short) was a small interferometer constructed over a series of expeditions to The Forks, Maine. By our last expedition in the summer of 2013, we deployed 64 dual-polarization MWA dipoles, all fully correlated. The purpose of the experiment was to demonstrate technology for highly scalable interferometers that use redundant calibration [124] which makes Fast Fourier Transform correlation possible [210]. More details on the design, deployment, and initial results from MITEoR can be found in Chapter 6.

1.3.4 Next Generation 21 cm Interferometers

While the first generation of 21 cm observatories is still taking and analyzing data, hoping to make a detection of the 21 cm signal, none can do much better than that. To not just detect but also characterize the power spectrum during the epoch of reionization, much larger telescopes are needed. Two are planned, the Hydrogen Epoch of Reionization Array and the Square Kilometre Array, each with different technological heritages and design philosophies.
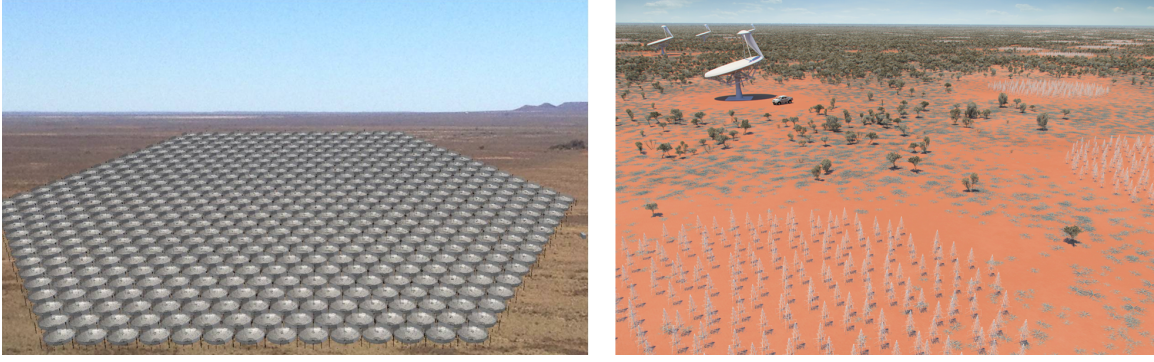
Figure 1-11: Renderings of next generation interferometers: the Hydrogen Epoch of Reionization Array (HERA, left) and the Square Kilometre Array (SKA, right). *Image credits: David DeBoer and the SKA Collaboration, respectively.*

1.3.4.1 The Hydrogen Epoch of Reionization Array

The Hydrogen Epoch of Reionization Array (HERA) is a planned focused EoR experiment. It will contain 352 crossed dipoles suspended at prime focus over fixed 14 m parabolic dishes (see the left panel of Figure 1-11). HERA is thus a pure drift-scanning instrument. The inner 331 dishes, which are constructed from telephone poles, wire mesh, and PVC pipe, are in a maximally packed hexagonal configuration.[3] HERA is funded under the NSF's Mid-Scale Innovations Program to begin construction with 37 dishes. Observations with the first 19 elements are scheduled to begin later this year.

HERA is the spiritual successor to PAPER; it has a densely packed, highly redundant configuration, a simple element design, and is being constructed on the PAPER site in South Africa. It maximizes the collecting area inexpensively by sacrificing sky coverage and the ability to point. Unlike PAPER, its Fourier sampling is dense enough that low-resolution, high-sensitivity imaging should be possible. While HERA is optimized for foreground avoidance, it may be possible to improve its performance with foreground subtraction. HERA's simple design will make this easier, though by no means easy.

Chapter 3 of this thesis was written with HERA in mind and uses HERA as a

---

[3]The hexagonal packing was my first and certainly my most visible contribution to the HERA design.

fiducial array. Chapter 8 is an analysis of the ability of HERA detect the EoR and constrain reionization parameters, though it was performed with an earlier design of HERA that called for 547 dishes in the hexagonal core. Regardless, a single observation season with HERA can definitely yield a robust EoR detection and scientifically novel constraints on the physics behind the EoR, even in the foreground avoidance regime.

1.3.4.2 The Square Kilometre Array

By contrast, the Square Kilometre Array (SKA) is the spiritual successor to LOFAR and, to a lesser extent, the MWA. The first phase (SKA1) of the long-planned telescope will actually be two telescopes, the SKA1-LOW near the MWA site in Australia and the SKA1-MID near the PAPER site in South Africa. The SKA1-LOW, the telescope relevant to 21 cm cosmology during the Cosmic Dawn will consist of 130,000 "christmas tree" dipoles (see the righthand panel of Figure 1-11) arranged into approximately 500 stations for a total collecting area of about $0.4 \, \mathrm{km}^2$. Each dipole will be individually digitized and station dipoles will be added together to form 30 simultaneous beams, each roughly 1 square degree. Construction of SKA1 is projected to begin in 2018 and be finished by 2023.

Unlike HERA, the SKA is a general purpose observatory with many different scientific objectives. Still, exploring the Cosmic Dawn via a number of probes is one key science drivers [130, 186, 4]. Like LOFAR, the SKA will have many fewer short baselines and much less redundancy than HERA, making redundant calibration and foreground avoidance more difficult. For that reason, despite its much larger collecting area the SKA's sensitivity will only be marginally better than HERA's if foregrounds can't be subtracted (see Greig and Mesinger [82] or Chapter 8 for estimates). On the other hand, with its increased collecting area and resolution, the SKA should be able to easily image the ionized bubbles [242], making it a more capable instrument for moving beyond the 21 cm power spectrum toward other statistical measurements of the EoR [110].

1.3.4.3 Omniscopes

The cost to build very large radio interferometers is eventually dominated by the cost of the correlator. Correlating every element with every other element usually requires computing resources that scale as N^2, where N is the number of antenna elements. The MWA, LOFAR, and the SKA avoid this problem using phased arrays, thereby not correlating every antenna with every other antenna, but rather tiles or groups of tiles together. GMRT and HERA get their sensitivity from large individual elements, instead of large N, at the cost of field of view. Eventually, if we want to very precisely test ΛCDM with 21 cm tomography, we'll want telescopes with both large fields of view and large collecting area [135]. The only way I know to achieve that is to build an interferometer that uses fast Fourier transform correlation.

All telescopes are Fourier transformers. Optical telescopes convert incoming photon momenta into positions in the focal plane. Interferometers sample the incoming radiation field in Fourier space using correlators to compare antenna signals at different baseline separations, effectively performing a discrete Fourier transform. It is not so surprising that, as Tegmark and Zaldarriaga [210] prove, any regular grid of antennas can be correlated with the fast Fourier transform (FFT). In fact, Tegmark and Zaldarriaga [211] showed that any hierarchically regular arrangement of elements can correlated with only $\mathcal{O}(N \log N)$ calculations and called that class of telescopes "omniscopes" for their broad spectral coverage and wide field of view.

Building such an array will be a major challenge. By design, they only save data from unique baselines, meaning that they must be calibrated in real time. Part of the motivation for Chapter 6 was to show that the technical advances necessary for this sort of telescope are within reach. HERA, with its highly redundant configuration, will also be an interesting testbed for FFT correlation. I believe that these designs are the future for 21 cm interferometers with truly massive collecting areas and I'm excited for what that future holds.

1.4 A Roadmap for this Thesis

The work that constitutes this thesis was originally written as seven different papers. The papers appear here as Chapters 2 through 8 and are reproduced verbatim with the permission of their primary co-authors. I played a significant role in the development and writing of all seven papers and served as the first author on four of them—in this thesis, Chapters 2, 3, 4, and 5. Six of them have already been published in peer-reviewed journals; Chapter 5 has been submitted and is still under review.

Instead of presenting the papers chronologically, I have organized this thesis into three thematic parts. In Part I, *Novel Data Analysis Tools*, I begin with two chapters devoted to rigorous but fast techniques for data analysis for 21 cm tomography.

- Chapter 2 reproduces the published paper *A fast method for power spectrum and foreground analysis for 21 cm cosmology* [58], written in collaboration with Adrian Liu and Max Tegmark. It presents a method for fast power spectrum estimation that extends and accelerates the method developed by Liu and Tegmark [120]. It also serves as a starting point for the rest of this thesis, much of which focuses on applying and refining these analysis techniques. The work in this chapter was conducted under the supervision of Max Tegmark in close consultation with Adrian Liu, but the project was lead and carried out largely by me.

- Chapter 3 reproduces the published paper *Mapmaking for precision 21 cm cosmology* [61], written in collaboration with Max Tegmark, Adrian Liu, Aaron Ewall-Wice, Jackie Hewitt, Miguel Morales, Abraham Neben, Aaron Parsons, and Jeff Zheng. It focuses on relaxing a key assumption in Chapter 2 that the PSF is not direction dependent. Understanding the precise statistics of interferometric maps is essential to the separation of Fourier space into the "EoR window" and the "wedge." Relaxing this assumption presents a number of computational difficulties, which the second half of the paper focuses on overcoming with a few well-controlled approximations. The work in this chapter was conducted under the supervision of Max Tegmark, whose appendix in Tegmark and

Zaldarriaga [211] served as the original inspiration for the paper, but was lead and carried out largely by me.

In Part II, *Early Results from New Telescopes*, I turn from the theoretical development of new data analysis techniques to the application of those methods (and related techniques developed in the Tegmark group) to real data from new radio telescopes—the Murchison Widefield Array and MITEoR. All three chapters in Part II refine previously published analysis techniques to help them meet the challenges of real data. Likewise, all three present the results of those analyses on early data from those telescopes.

- Chapter 4 reproduces the published paper *Overcoming real-world obstacles in 21 cm power spectrum estimation: A method demonstration and results from early Murchison Widefield Array data* [59], co-authored with Adrian Liu and written in collaboration with Chris Williams, Jackie Hewitt, Max Tegmark, and a number of other MWA members. It discusses numerous challenges presented by real-world data that the idealized analyses of Liu and Tegmark [120] and Chapter 2 ignored or glossed-over and found ways to consistently deal with them in order to produce the MWA's first limit on the 21 cm power spectrum. The paper was an equal effort by Adrian Liu and myself. Adrian developed the majority of the methods detailed in Section 4.2 and wrote most of that section. The data was prepared by Chris Williams and I performed the method demonstration and power spectrum analysis that constituted Section 4.3, most of which I wrote.

- Chapter 5 reproduces the paper *Empirical covariance modeling for 21 cm power spectrum estimation: A method demonstration and new limits from early Murchison Widefield Array 128-tile data* [60] which is currently being reviewed by *Physical Review D*. It was written in collaboration with Abraham Neben and under the supervision of Jackie Hewitt and Max Tegmark; the MWA EoR collaboration and Builder's List are also co-authors. The paper is a follow-up to Chapter 4 and similarly presents new limits on the 21 cm power spectrum with

a few hours of MWA observations. These limits demonstrate the efficacy of the method developed to estimate realistic foreground residual covariance models that are empirically motivated but constrained by our prior beliefs about the frequency structure of the foregrounds. Abraham prepared the maps for power spectrum analysis, provided some of the original ideas for covariance estimation in Fourier space, and wrote Sections 5.3.1 and 5.3.2. I developed the empirical covariance estimation method, performed the power spectrum analysis, and wrote the rest of the paper.

- Chapter 6 reproduces the published paper *MITEoR: a scalable interferometer for precision 21 cm cosmology* [248], authored by Jeff Zheng under the supervision of Max Tegmark. Jeff performed the plurality of the work bringing the MITEoR projection to fruition, though it was the culmination of years of effort in the Tegmark group to build and demonstrate an interferometer capable of real-time FFT correlation. I am the fourth author on the paper. My role in the project varied over the years and included data analysis for the first expedition, deployment of several later expeditions, satellite tracking software, and visibility simulations. While Jeff performed the final data analysis and wrote the majority of this paper, I served in a consulting role during the development of the techniques discussed and, along with Max and Adrian Liu, as the primary editor of the paper. Many undergraduate researchers, graduate students, postdocs, and other scientists contributed to the MITEoR project and are also authors on the paper.

Finally, in Part III, *The Cosmic Dawn on the Horizon*, I look forward to what we might be able to measure with next generation 21 cm interferometers. This part includes two chapters based on previously published forecasts that examine the potential for astrophysical constraints on the first stars, galaxies, black holes and their effect on the IGM.

- Chapter 7 reproduces the published paper *Detecting the 21 cm forest in the 21 cm power spectrum* [63], written with Aaron Ewall-Wice, Andrei Mesinger,

and Jackie Hewitt. The paper investigates the effect of 21 cm absorption along lines of sight to high-redshift, radio-loud quasars on the the 21 cm power spectrum. While the effect depends on the relatively unconstrained population of high-redshift quasars and on the thermal history of the IGM, it potentially has a detectible and distinguishable impact on future measurements. This project was lead by Aaron, who performed most of the analysis and wrote most of the paper. Andrei performed the IGM simulations and Jackie supervised the project. As second author, I performed the detailed detectability calculations in Section 7.5 and served as the primary editor.

- Chapter 8 reproduces the published paper *What Next-Generation 21 cm Power Spectrum Measurements Can Teach Us About the Epoch of Reionization* [184], written by Jonnie Pober, Adrian Liu, and myself in collaboration with several other members of the HERA team. The work began with a detailed sensitivity calculation comparison between Jonnie and myself, which eventually led the calculation of the errors HERA should expect on a measurement of the power spectrum for a variety of reionization models and foreground mitigation strategies (Section 8.3). Adrian followed up that work with a detailed Fisher matrix analysis of the potential constraints on a parameterized model of reionization in Section 8.3.5.

It is my hope that this thesis presents a broad picture of how we might eventually overcome the difficulties of detecting the 21 cm signal, the progress we have already made with the first generation of telescopes, and the exciting science we'll be able to do with those measurements.

Part I

Novel Data Analysis Tools

Chapter 2

A Fast Method for Power Spectrum and Foreground Analysis for 21 cm Cosmology

The content of this chapter was submitted to Physical Review D *on November 27, 2012 and published [58] as* A fast method for power spectrum and foreground analysis for 21 cm cosmology *on February 12, 2013.*

2.1 Introduction

Neutral hydrogen tomography with the 21 cm line promises to shed light on vast and unexplored epoch of the early universe. As a cosmological probe, it offers the opportunity to directly learn about the evolution of structure in our universe during the cosmological dark ages and the subsequent Epoch of Reionization (EoR) [11, 74, 154, 188]. More importantly, the huge volume of space and wide range of cosmological scales probed makes 21 cm tomography uniquely suited for precise statistical determination of the parameters that govern modern cosmological and astrophysical models for how our universe transitioned from hot and smooth to cool and clumpy [139, 194, 26, 75, 241, 38, 134, 222, 221, 98, 71, 128, 72, 12, 131, 135, 44]. It has the potential to surpass even the Cosmic Microwave Background (CMB) in its sensitivity

as a cosmological probe [135].

The central idea behind $21\,\mathrm{cm}$ tomography is that images produced by low frequency radio interferometers at different frequencies can create a series of images at different redshifts, forming a three dimensional map of the $21\,\mathrm{cm}$ brightness temperature. Yet we expect that our images will be dominated by synchrotron emission from our galaxies and others. In fact, we expect those foreground signals to dominate over the elusive cosmological signal by about four orders of magnitude [53, 19].

One major challenge for $21\,\mathrm{cm}$ cosmology is the extraction of the brightness temperature power spectrum, a key prediction of theoretical models of the dark ages and the EoR, out from underneath a mountain of foregrounds and instrumental noise. Liu & Tegmark ([120], hereafter "LT") presented a method for power spectrum estimation that has many advantages over previous approaches (on which we will elaborate in Section 2.2). It has, however, one unfortunate drawback: it is very slow. The LT method relies on multiplying and inverting very large matrices, operations that scale as $\mathcal{O}(N^3)$, where N is the number of voxels of data to analyze.

The goal of the present paper is to develop and demonstrate a way of achieving the results of the LT method that scales only as $\mathcal{O}(N \log N)$. Along the way, we will also show how LT can be extended to take advantage of additional information about the brightest point sources in the map while maintaining a reasonable algorithmic scaling with N. Current generation interferometers, including the Low Frequency Array (LOFAR, [76]), the Giant Metrewave Radio Telescope (GMRT, [166]), the Murchinson Widefield Array (MWA, [220]), and the Precision Array for Probing the Epoch of Reionization (PAPER, [171]) are already producing massive data sets at or near the megavoxel scale (e.g. [234]). These data sets are simply too large to be tackled by the LT method. We expect next generation observational efforts, like the Hydrogen Epoch of Reionization Array [81], a massive Omniscope [211], or the Square Kilometer Array [33], to produce even larger volumes of data. Moreover, as computer processing speed continues to grow exponentially, the ability to observe with increasingly fine frequency resolution will enable the investigation of the higher Fourier modes of the power spectrum at the cost of yet larger data sets. The need

for an acceleration of the LT method is pressing and becoming more urgent.

Our paper has a similar objective to [175], which also seeks to speed up algorithms for power spectrum estimation with iterative and Monte Carlo techniques. The major differences between the paper arise from the our specialization to the problem of 21 cm cosmology and the added complications presented by foregrounds, especially with regard to the basis in which various covariance matrices are easiest to manipulate. Our paper also shares similarities to [205]. Like [175], [205] does not extend its analysis to include foregrounds. It differs also from this paper in spirit because that it seeks to go from interferometric visibilities to a power spectrum within a Bayesian framework rather than from a map to a power spectrum and because it considers one frequency channel at a time. In this paper, we take advantage of many frequency channels simultaneously in order to address the problem of foregrounds.

This paper is organized as follows. We begin with Section 2.2 wherein we review the motivation for and details of the LT method. In Section 2.3 we present the novel aspects of our technique for measuring the 21 cm brightness temperature power spectrum. We discuss the extension of the method to bright point sources and the assumptions we must make to accelerate our analysis. In Section 2.4 we demonstrate end-to-end tests of the algorithm and show some of its first predictions for the ability of the upcoming 128-tile deployment of the MWA to detect the statistical signal of the Epoch of Reionization.

2.2   The Brute Force Method

The solution to the problem of power spectrum estimation in the presence of foregrounds put forward by LT offers a number of improvements over previous proposals that rely primarily on line of sight foreground information [231, 78, 28, 123, 107, 86, 122, 87, 41]. The problem of 21 cm power spectrum estimation shares essential qualities with both CMB and galaxy survey power spectrum estimation efforts. Like with galaxy surveys, we are interested in measuring a three dimensional power spectrum. On the other hand, our noise and foreground contaminants bear more similarity to

the problems faced by CMB studies—though the foregrounds we face are orders of magnitude larger.

The LT method therefore builds on the literature of both CMB and galaxy surveys, providing a unified framework for the treatment of geometric and foreground effects by employing the quadratic estimator formalism for inverse variance foreground subtraction and power spectrum estimation. In Section 2.2.3 we will review precisely how it is implemented.

The LT formalism has a number of important advantages over its predecessors. By treating foregrounds as a form of correlated noise, both foregrounds and noise can be downweighted in a way that is unbiased and lossless in the sense that it maintains all the cosmological information in the data. Furthermore, the method allows for the simultaneous estimation of both the errors on power spectrum estimates and the window functions or "horizontal" error bars.

Unfortunately, the LT method suffers from computational difficulties. Because it involves inverting and multiplying very large matrices, it cannot be accomplished faster than in $\mathcal{O}(N^3)$ steps, where N is the number of voxels in the data to be analyzed. This makes analyzing large data sets with this method infeasible. The primary goal of this paper is to demonstrate an adaptation of the method that can be run much faster. But first, we need to review the essential elements of the method to put our adaptations and improvements into proper context. In Sections 2.2.1 and 2.2.2, we describe our conventions and notation and explain the relationship between the measured quantities and those we seek to estimate. In Section 2.2.3, we review the LT statistical estimators and how the Fisher information matrix is used to calculate statistical errors on power spectrum measurements. Then in 2.2.4 we explain the LT model of noise and foregrounds in order to motivate and justify our refinements that will greatly speed up the algorithm in Section 2.3.

2.2.1 Data Organization and Conventions

We begin with a grid of data that represents the brightness temperatures at different positions on the sky as a function of frequency from which we wish to estimate
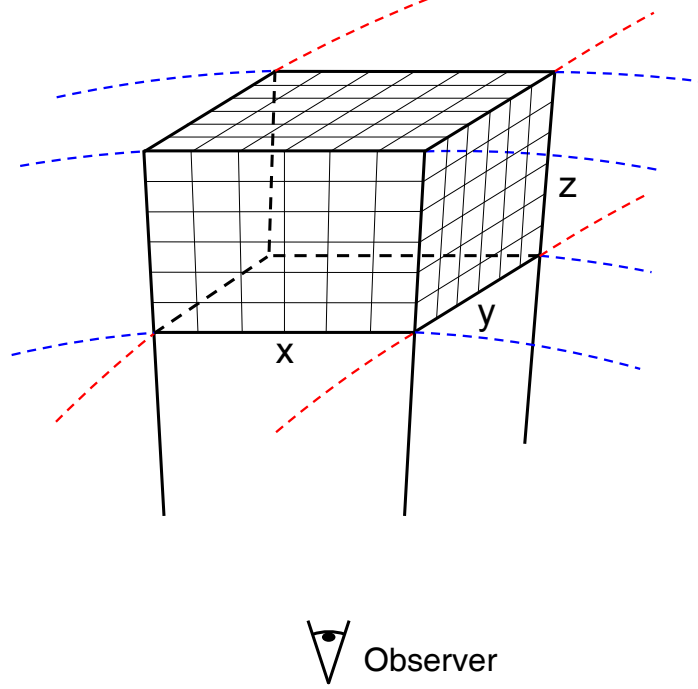
Figure 2-1: This exaggerated schematic illustrates the flat sky approximation. It shows great circles (colored and dashed) approximated linearly in the region considered, with lines tracing back to the observer treated as if they were parallel. Our data cube contains the measured brightness temperatures for every small voxel.

the 21 cm brightness temperature power spectrum. We summarize that information using a data vector $\mathbf{x}$ which can be thought of as a one dimensional object of length $n_x n_y n_z \equiv N$,[1] the number of voxels in the data cube.

Although the LT technique works for arbitrary survey geometries, we restrict ourselves to the simpler case of a data "cube" that corresponds to a relatively small rectilinear section of our universe of size $\ell_x \times \ell_y \times \ell_z$ in comoving coordinates.[2] We pick our box to be a subset of the total 21 cm brightness temperature 3D map that a large interferometric observatory would produce. Unlike the LT method, our fast method requires that the range of positions on the sky must be small enough for the flat sky approximation to hold (Figure 2-1). Similarly, our range of frequencies (and thus redshifts) in the data cube must correspond to an epoch short enough so that

---

[1] While it is helpful to think of $\mathbf{x}$ as a vector in the matrix operations below, it is important to remember that the index i in x_i, which refers to the different components of $\mathbf{x}$, actually runs over different values of the spatial coordinates x, y, and z.

[2] This restriction and its attendant approximations lie at the heart of our strategy for speeding up these calculations, as we explain in Section 2.3.

$P(k, z)$ might be approximated as constant in time. Following simulations by [140], [135] argued that we can conservatively extend the redshift ranges of our data cubes to about $\Delta z \lesssim 0.5$. At typical EoR redshifts, such a small range in Δz allows a very nearly linear mapping between the frequencies measured by an interferometer to a regularly spaced range of comoving distances, $d_C(z)$, although in general $d_C(z)$ is not a linear function of z or ν. This also justifies the approximation that our data cube corresponds to an evenly partitioned volume of our universe.

If the measured brightness temperatures, x_i, were only the result of redshifted 21 cm radiation, then each measurement would represent the average value in some small box of volume $\Delta x \Delta y \Delta z$ centered on $\mathbf{r}_i$ of a continuous brightness temperature field $x(\mathbf{r})$ [213]:

$$x_i \equiv \int \psi_i(\mathbf{r}) x(\mathbf{r}) d^3 r, \tag{2.1}$$

where our discretization function ψ_i is defined as $\psi_i(\mathbf{r}) \equiv \psi_0(\mathbf{r} - \mathbf{r}_i)$, where

$$\psi_0(\mathbf{r}) \equiv \frac{\Pi(\frac{x}{\Delta x})\Pi(\frac{y}{\Delta y})\Pi(\frac{z}{\Delta z})}{\Delta x \Delta y \Delta z} \tag{2.2}$$

and where $\Pi(x)$ is the normalized, symmetric boxcar function ($\Pi(x) = 1$ if $|x| < \frac{1}{2}$ and 0 otherwise). This choice of pixelization encapsulates the idea that each measured brightness temperature is the average over a continuous temperature field inside each voxel. In this paper, we improve on the LT method by including the effect of finite pixelation. This will manifest itself as an extra $\Phi(\mathbf{k})$ term that will we define in Equation 2.6 and that will reappear throughout this paper.

2.2.2 The Discretized 21 cm Power Spectrum

Ultimately, the goal of this paper is to estimate the 21 cm power spectrum $P(\mathbf{k})$, defined via

$$\langle \widetilde{x}^*(\mathbf{k})\widetilde{x}(\mathbf{k}')\rangle \equiv (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}')P(\mathbf{k}), \tag{2.3}$$

where $\widetilde{x}(\mathbf{k})$ is the Fourier transformed brightness temperature field and where angle brackets denote the ensemble average of all possible universes obeying the same

statistics.

Our choice of pixelization determines the relationship between the continuous power spectrum, $P(\mathbf{k})$, and the 21 cm signal covariance matrix, which we call $\mathbf{S}$ for **Signal**. It is fairly straightforward to show, given Equation 2.1 and the definition of the power spectrum [213], that:

$$S_{ij} \equiv \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \int \widetilde{\psi}_i(\mathbf{k}) \widetilde{\psi}_j^*(\mathbf{k}) P(\mathbf{k}) \frac{d^3 k}{(2\pi)^3}, \tag{2.4}$$

where $\widetilde{\psi}_i(\mathbf{k})$ is the Fourier transform of $\psi_i(\mathbf{r})$:

$$\widetilde{\psi}_i(\mathbf{k}) \equiv \int e^{-i\mathbf{k}\cdot\mathbf{r}} \psi_i(\mathbf{r}) d^3\mathbf{r}. \tag{2.5}$$

Separating this integral into each of the three Cartesian coordinates and integrating yields

$$\widetilde{\psi}_i(\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{r}_i} \Phi(\mathbf{k}), \text{ where}$$
$$\Phi(\mathbf{k}) \equiv j_0\left(\frac{k_x \Delta x}{2}\right) j_0\left(\frac{k_y \Delta y}{2}\right) j_0\left(\frac{k_z \Delta z}{2}\right), \tag{2.6}$$

where $j_0(x) = \sin x / x$ is the zeroth spherical Bessel function. Because we can only make a finite number of measurements of the power spectrum, we parametrize and discretize $P(\mathbf{k})$ by approximating it as a piecewise constant function:

$$P(\mathbf{k}) \approx \sum_\alpha p^\alpha \chi^\alpha(\mathbf{k}), \tag{2.7}$$

where the "band power" p^α gives the power in region α of Fourier space,[3] specified by the characteristic function $\chi^\alpha(\mathbf{k})$ which equals 1 inside the region and vanishes elsewhere.

---

[3]In contrast to lowered Latin indices, which we use to pick out voxels in a real space or Fourier space data cube, we will use raised Greek indices to pick out power spectrum bins, which will generally each run a range in $k_\parallel$ and in $k_\perp$.

Combining Equations 2.4 and 2.7 we can write down S_{ij}:

$$S_{ij} = \sum_\alpha p^\alpha Q_{ij}^\alpha, \text{ where}$$

$$Q_{ij}^\alpha \equiv \int \widetilde{\psi}_i(\mathbf{k}) \widetilde{\psi}_j^*(\mathbf{k}) \chi^\alpha(\mathbf{k}) \frac{d^3k}{(2\pi)^3}. \tag{2.8}$$

We choose these $\chi^\alpha(\mathbf{k})$ to produce band powers that reflect the symmetries of the observation. Our universe is isotropic in three dimensions, but due to redshift space distortions, foregrounds, and other effects, our measurements will be isotropic only perpendicular to the line of sight [135, 13, 5, 161, 9]. This suggests cylindrical binning of the power spectrum; in the directions perpendicular to the line of sight, we bin k_x and k_y together radially to get a region in k-space extending from from $k_\perp^\alpha - \Delta k_\perp/2$ to $k_\perp^\alpha + \Delta k_\perp/2$ where $k_\perp^2 \equiv k_x^2 + k_y^2$. Likewise, in the direction parallel to the line of sight, we integrate over a region of k-space both from $k_\parallel^\alpha - \Delta k_\parallel/2$ to $k_\parallel^\alpha + \Delta k_\parallel/2$ and, because the power spectrum only depends on $k_\parallel \equiv |k_z|$, from $-k_\parallel^\alpha + \Delta k_\parallel/2$ to $-k_\parallel^\alpha - \Delta k_\parallel/2$. Therefore, we have

$$Q_{ij}^\alpha = \frac{1}{(2\pi)^3} \left[\int_{k_\parallel^\alpha - \Delta k_\parallel/2}^{k_\parallel^\alpha + \Delta k_\parallel/2} - \int_{-k_\parallel^\alpha + \Delta k_\parallel/2}^{-k_\parallel^\alpha - \Delta k_\parallel/2} \right] \int_{k_\perp^\alpha - \Delta k_\perp/2}^{k_\perp^\alpha + \Delta k_\perp/2} |\Phi(\mathbf{k})|^2 e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} k_\perp \, d\theta dk_\perp dk_\parallel.$$

$$\tag{2.9}$$

Without the factor of $|\Phi(\mathbf{k})|^2$, the LT method was able to evaluate this integral analytically. With it, the integral must be evaluated numerically if it is to be evaluated at all. This is of no consequence; we will return to this formula in Section 2.3.2 to show how the matrix $\mathbf{Q}^\alpha$ naturally lends itself to approximate multiplication by vectors using fast Fourier techniques.

2.2.3   21 cm Power Spectrum Statistics

In order to interpret the data from any experiment, we need to be able to estimate both the 21 cm brightness temperature power spectrum and the correlated errors induced by the survey parameters, the instrument, and the foregrounds. The LT method

does both at the same time; with it, the calculation of the error bars immediately enables power spectrum estimation.

2.2.3.1 Inverse Variance Weighted Power Spectrum Estimation

The LT method adapts the inverse variance weighted quadratic estimator formalism [208, 23] for calculating 21 cm power spectrum statistics. The first step towards constructing the estimator $\widehat{p}^\alpha$ for p^α is to compute a quadratic quantity, called $\widehat{q}^\alpha$ whose relationship[4] to $\widehat{p}^\alpha$ we will explain shortly:

$$\widehat{q}^\alpha \equiv \frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^{\mathsf{T}} \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle). \tag{2.10}$$

Here $\mathbf{C}$ is the covariance matrix of $\mathbf{x}$, so

$$\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^{\mathsf{T}} \rangle - \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^{\mathsf{T}}. \tag{2.11}$$

For any given value of α, the right-hand side of Equation (2.10) yields a scalar. Were both our signal and foregrounds Gaussian, this estimator would be optimal in the sense that it preserves all the cosmological information contained in the data. Of course, with a non-Gaussian signal, the power spectrum cannot contain all of the information, though it still can be very useful [208].

Our interest in the quadratic estimators $\widehat{q}^\alpha$ lies in their simple relationship to the underlying band powers. In [208], it is shown that:

$$\langle \widehat{\mathbf{q}} \rangle = \mathbf{F}\mathbf{p} + \mathbf{b} \tag{2.12}$$

where each b^α is the bias in the estimator and $\mathbf{F}$ is the Fisher information matrix, which is related to the probability of having measured our data given a particular set

---

[4]Unlike the notation in LT, we do not include the bias term in $\widehat{q}^\alpha$ but will later include it in our power spectrum estimator. The result is the same.

of band powers, $f(\mathbf{x}|p^\alpha)$. The matrix is defined [67] as:

$$F^{\alpha\beta} \equiv -\left\langle \frac{\partial^2 \ln f(\mathbf{x}|p^\alpha)}{\partial p^\alpha \partial p^\beta} \right\rangle. \tag{2.13}$$

The LT method employs the estimators by calculating both $\mathbf{F}$ and $\mathbf{b}$ using relationships derived in [208]:

$$F^{\alpha\beta} = \frac{1}{2}\text{tr}[\mathbf{C}^{-1}\mathbf{Q}^\alpha\mathbf{C}^{-1}\mathbf{Q}^\beta] \text{ and} \tag{2.14}$$

$$b^\alpha = \frac{1}{2}\text{tr}[(\mathbf{C}-\mathbf{S})\mathbf{C}^{-1}\mathbf{Q}^\alpha\mathbf{C}^{-1}]. \tag{2.15}$$

We want our $\widehat{p}^\alpha$ to be unbiased estimators of the true underlying band powers, which means that we will have to take care to remove the biases for each band power, b^α. We construct our estimators[5] as linear combinations of the quadratic estimators $\widehat{q}^\alpha$ that have been corrected for bias:

$$\widehat{\mathbf{p}} = \mathbf{M}(\widehat{\mathbf{q}} - \mathbf{b}), \tag{2.16}$$

where $\mathbf{M}$ is a matrix which correctly normalizes the power spectrum estimates; the form of $\mathbf{M}$ represents a choice in the trade-off between small error bars and narrow window functions, as we will explain shortly.

How do we expect this estimator to behave statistically? The only random variable on the right hand side of Equation 2.16 is $\widehat{\mathbf{q}}$, so we can combine Equations 2.12 and 2.16 to see that our choice of $\widehat{\mathbf{p}}$ indeed removes the bias term:

$$\langle \widehat{\mathbf{p}} \rangle = \mathbf{M}\mathbf{F}\mathbf{p} + \mathbf{M}\mathbf{b} - \mathbf{M}\mathbf{b} = \mathbf{M}\mathbf{F}\mathbf{p} = \mathbf{W}\mathbf{p}. \tag{2.17}$$

We have defined the matrix of "window functions" $\mathbf{W} \equiv \mathbf{M}\mathbf{F}$ because Equation 2.17 tells us that we can expect our band power spectrum estimator, $\widehat{\mathbf{p}}$, be be a weighted average of the true, underlying band powers, $\mathbf{p}$. That definition imposes the condition

---

[5]Here were differ slightly from the LT method in the normalization, which does not have the property from Equation 2.18. We instead follow [214].

on $\mathbf{W}$ that

$$\sum_\beta W^{\alpha\beta} = 1 \tag{2.18}$$

which is equivalent to the statement that the weights in a weighted average must add up to one. The condition on $\mathbf{W}$ constrains our choice of $\mathbf{M}$, though as long as $\mathbf{M}$ is an invertible matrix,[6] the choice of $\mathbf{M}$ does not change the information content of our power spectrum estimate, only the way we choose to represent our result.

2.2.3.2 Window Functions and Error Bars

In this paper, we choose a form of $\widehat{\mathbf{p}}$ where $\mathbf{M} \propto \mathbf{F}^{-1/2}$. Two other choices for $\mathbf{M}$ are presented in [214]: one where $\mathbf{M} \propto \mathbf{I}$ and another where $\mathbf{M} \propto \mathbf{F}^{-1}$. The former produces the smallest possible error bars, but at the cost of wide window functions and correlated measurement errors. The latter produces δ-function windows, but large and anticorrelated measurement errors. This choice of $\mathbf{M} \propto \mathbf{F}^{-1/2}$ has proven to be a happy medium between those other two choices for $\mathbf{M}$. It produces reasonably narrow window functions and reasonably small error bars which have the added advantage of being completely uncorrelated, so that each measurement contains a statistically independent piece of information. Because $\mathbf{W} \equiv \mathbf{MF}$ and because of the condition on $\mathbf{W}$ in Equation 2.18, there is only one such $\mathbf{M}$:

$$M^{\alpha\beta} \equiv \frac{\left(\mathbf{F}^{-1/2}\right)^{\alpha\beta}}{\sum_\gamma (\mathbf{F}^{1/2})^{\alpha\gamma}}. \tag{2.19}$$

With this choice of $\mathbf{M}$ we get window functions of the form

$$W^{\alpha\beta} = \frac{(\mathbf{F}^{1/2})^{\alpha\beta}}{\sum_\gamma (\mathbf{F}^{1/2})^{\alpha\gamma}} \tag{2.20}$$

which we can use to put "horizontal error bars" on our power spectrum estimates.

Using Equation 2.16 and the fact derived in [208] that an equivalent formula for

---

[6]None of the choices of $\mathbf{M}$ involve anything more computationally intensive than inverting $\mathbf{F}$. This is fine, since $\mathbf{F}$ is a much smaller matrix than $\mathbf{C}$.

F is given by

$$\mathbf{F} = \langle \widehat{\mathbf{q}}\widehat{\mathbf{q}}^{\mathsf{T}} \rangle - \langle \widehat{\mathbf{q}} \rangle \langle \widehat{\mathbf{q}} \rangle^{\mathsf{T}}, \tag{2.21}$$

we can see that the covariance of $\widehat{\mathbf{p}}$ takes on a simple form:

$$\langle \widehat{\mathbf{p}}\widehat{\mathbf{p}}^{\mathsf{T}} \rangle - \langle \widehat{\mathbf{p}} \rangle \langle \widehat{\mathbf{p}} \rangle^{\mathsf{T}} = \mathbf{MFM}^{\mathsf{T}}. \tag{2.22}$$

This allows us to write down the "vertical error bars" on our individual power spectrum estimates:

$$\Delta \widehat{p}^{\alpha} = \left[\left(\mathbf{MFM}^{\mathsf{T}} \right)^{\alpha\alpha} \right]^{1/2} = \frac{1}{\sum_{\gamma} (\mathbf{F}^{1/2})^{\alpha\gamma}}. \tag{2.23}$$

As in LT, we can transform our power spectrum estimates and our vertical error bars into temperature units:

$$\widehat{T}^{\alpha} \equiv \left[\frac{(k_{\perp}^{\alpha})^2 k_{\parallel}^{\alpha}}{2\pi^2} p^{\alpha} \right]^{1/2} \tag{2.24}$$

and likewise,

$$\Delta \widehat{T}^{\alpha} = \left[\frac{(k_{\perp}^{\alpha})^2 k_{\parallel}^{\alpha}}{2\pi^2 \left(\sum_{\gamma} (\mathbf{F}^{1/2})^{\alpha\gamma} \right)} \right]^{1/2}. \tag{2.25}$$

This makes it easier to compare to theoretical predictions, which are often quoted in units of K or mK.

2.2.4 Foreground and Noise Models

The structure of the matrix **C** that goes into our inverse variance weighted estimator depends on the way we model our foregrounds, noise, and signal. We assume that those contributions are the sum of five uncorrelated components:

$$\mathbf{C} = \sum_{c\,\in\,\text{components}} \langle \mathbf{x}_c \mathbf{x}_c^{\mathsf{T}} \rangle - \langle \mathbf{x}_c \rangle \langle \mathbf{x}_c \rangle^{\mathsf{T}}$$

$$\equiv \mathbf{S} + \mathbf{R} + \mathbf{U} + \mathbf{G} + \mathbf{N}. \tag{2.26}$$

These are the covariance matrices due to 21 cm **S**ignal, bright point sources **R**esolved from one another, **U**nresolved point sources, the **G**alactic synchrotron, and detector

| Resolved Point Sources | Unresolved Point Sources | Galactic Synchrotron | Detector Noise |
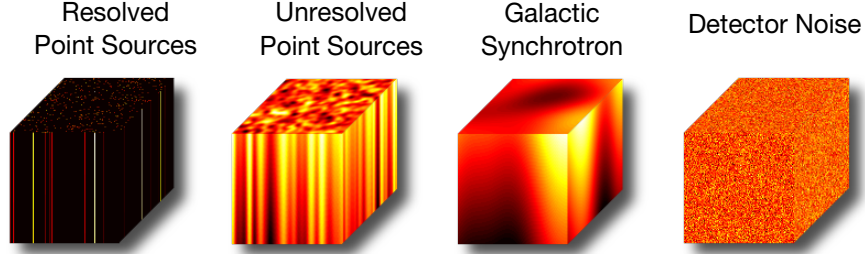
Figure 2-2: These example data cubes (with the line of sight drawn vertically) illustrate the strong or weak correlations between different voxels in the same cube. In Section 2.3.6 we explain how these simulated data cubes are generated quickly. The addition of resolved point sources, which is not included in LT, is discussed in Section 2.3.4.1. To best exemplify the detailed structure of the models, the color scales are different for each of the cubes.

Noise, respectively. This deconstruction of $\mathbf{C}$ is both physically motivated and will ultimately let us approximate $\mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle)$ much more quickly than by just inverting the matrix.

Following LT, we neglect the small cosmological $\mathbf{S}$ because it is only important for taking cosmic variance into account. It is straightforward to include the $\mathbf{S}$ matrix in our method, especially because we expect it to have a very simple form, but this will only be necessary once the experimental field moves from upper limits to detection and characterization of the 21 cm brightness temperature power spectrum.

In this paper, we will develop an accelerated version of the LT method using the models delineated in LT. That speed-up relies on the fact that all of these covariance matrices can be multiplied by vectors $\mathcal{O}(N\mathrm{log}N)$ time. However, our techniques for acceleration will work on a large class of models for $\mathbf{C}$ as long as certain assumptions about translation invariance and spectral structure are respected. In this section, we review the three contaminant matrices from LT: $\mathbf{U}$, $\mathbf{G}$, and $\mathbf{N}$. When we discuss methods to incorporate these matrices into a faster technique in Section 2.3.4, we will also expand the discussion of foregrounds to include $\mathbf{R}$, which is a natural extension of $\mathbf{U}$.

79

2.2.4.1 Unresolved Point Sources

For a typical current generation or near future experiment, the pixels perpendicular to the line of sight are so large that every one is virtually guaranteed to have a point source in it bright enough to be an important foreground to our 21 cm signal. These confusion limited point sources are taken into account using their strong correlations parallel to the line of sight and weaker correlations perpendicular to the line of sight, both of which are easily discerned in Figure 2-2.

Following LT we split $\mathbf{U}$ into the tensor product of two parts, one representing correlations perpendicular to the line of sight and the other parallel to the line of sight:

$$\mathbf{U} \equiv \mathbf{U}_\perp \otimes \mathbf{U}_\parallel \tag{2.27}$$

Covariance perpendicular to the line of sight is modeled as an unnormalized Gaussian:

$$(U_\perp)_{ij} \equiv \exp\left[\frac{((\mathbf{r}_\perp)_i - (\mathbf{r}_\perp)_j)^2}{2\sigma_\perp^2} \right] \tag{2.28}$$

where $\sigma_\perp$ represents the correlation length perpendicular to the line of sight. Following LT, we take this to be a comoving distance corresponding to 7 arcminutes, representing the weak clustering of point sources.

The covariance along the line of sight assumes a Poisson distributed number of point sources below some flux cut, S_{cut}, which we take to be 0.1 Jy, each with a spectral index drawn from a Gaussian distribution with mean $\bar{\kappa}$ and standard deviation σ_κ. Given a differential source count [57] of

$$\frac{dn}{dS} = (4000 \text{ Jy}^{-1}\text{sr}^{-1}) \times \begin{cases} \left(\frac{S}{0.880 \text{ Jy}}\right)^{-2.51} & \text{for S} > 0.880 \text{ Jy} \\ \left(\frac{S}{0.880 \text{ Jy}}\right)^{-1.75} & \text{for S} \le 0.880 \text{ Jy,} \end{cases} \tag{2.29}$$

we get a covariance parallel to the line of sight of

$$(U_\parallel)_{ij} = (1.4 \times 10^{-3} \text{ K})^2 \, (\eta_i\eta_j)^{-2-\bar{\kappa}} \left(\frac{\Omega_{pix}}{1 \text{ sr}}\right)^{-1} \exp\left[\frac{\sigma_\kappa^2}{2}(\ln(\eta_i\eta_j))^2\right] I_2(S_{cut}). \tag{2.30}$$

where we have assumed a power law spectrum for the point sources where $\eta_i \equiv \nu_i/\nu_*$, $\nu_* = 150$ MHz, and $\overline{\kappa}$ and σ_κ are the average value and standard deviation of the distribution of spectral indices of the point sources. We define $I_2(S_{cut})$ as

$$I_2(S_{cut}) \equiv \int_0^{S_{cut}} S^2 \frac{dn}{dS} dS \qquad (2.31)$$

Following LT, we take $\overline{\kappa} = 0.5$ and $\sigma_\kappa = 0.5$, both of which are consistent with the results of [234]. In Section 2.3.4.2, we will return to Equation 2.30 and show how it can be put into an approximate form that can be quickly multiplied by a vector.

2.2.4.2 Galactic Synchrotron Radiation

Following LT, we model Galactic synchrotron emission in the same way that we model unresolved point sources. Fundamentally, both are spatially correlated synchrotron signals contributing to the brightness temperature of every pixel in our data cube. However the galactic synchrotron is much more highly correlated spatially, which can be clearly seen in the sample data cube in Figure 2-2. This leads to our adoption of a much larger value of $\sigma_\perp$; we take $\sigma_\perp$ to be a comoving distance corresponding to $30°$ on the sky. Following LT, we take $\overline{\kappa} = 0.8$ and $\sigma_\kappa = 0.4$.

This is an admittedly crude model for the galactic synchrotron, in part because it fails to take into account the roughly planar spatial distribution of the Galactic synchrotron. A more sophisticated model for $\mathbf{G}$ that incorporates a more informative map of the Galactic synchrotron can only produce smaller error bars and narrower window functions. However, such a model might involve breaking the assumption of the translational invariance of correlations, which could be problematic for the technique we use in Section 2.3 to speed up this algorithm. In practice, we expect very little benefit from an improved spatial model of the Galactic synchrotron due to the restriction imposed by the flat sky approximation that our map encompass a relatively small solid angle.

2.2.4.3 Instrumental Noise

Here we diverge from LT to adopt a form of the noise power spectrum from [210] that is more readily adaptable to the pixelization scheme we will introduce:

$$P^N(\mathbf{k}, \lambda) = \frac{\lambda^2 T_{\text{sys}}^2 y d_M^2}{A f^{\text{cover}} \tau} B^{-2}(\mathbf{k}, \lambda). \tag{2.32}$$

Here T_{sys} is the system temperature (which is sky noise dominated in our case), A is the total effective collecting area of the array, and τ is the is the total observing time. $B(\mathbf{k}, \lambda)$ is a function representing the uv-coverage, normalized to peak at unity, which changes with wavelength. Lastly, y is the conversion from bandwidth to the comoving length of the box parallel to the line of sight and d_M is the transverse comoving distance[7], so $y d_M^2 \Omega_{\text{pix}} \Delta\nu = \Delta x \Delta y \Delta z$ with Ω_{pix} being the angular size of our pixels and $\Delta\nu$ being the frequency channel width. This form of the noise power spectrum assumes that the entire map is observed for the same time τ, which is why the ratio of the angular size of the map to the field of view does not appear.

We use Equation 2.4 to discretize the power spectrum and get $\mathbf{N}$:

$$N_{ij} = \int e^{i\mathbf{k}\cdot\mathbf{r}_i} e^{-i\mathbf{k}\cdot\mathbf{r}_j} |\Phi(\mathbf{k})|^2 P^N(\mathbf{k}, \lambda) \frac{d^3 k}{(2\pi)^3} \tag{2.33}$$

Instead of evaluating this integral, we will show in Section 2.3.4.3 that it can be approximated using the discrete Fourier transform.

2.2.5 Computational Challenges to the Brute Force Method

For a large data cube, the LT method requires the application of large matrices that are memory-intensive to store and computationally infeasible to invert. However, we need to be able to multiply by and often invert these large matrices to calculate our quadratic estimators (Equations 2.9 and 2.10), the Fisher matrix (Equation 2.14), and

---

[7]The transverse comoving distance, $d_M(z)$, is the ratio of an object's comoving size the angle it subtends, as opposed to the angular diameter distance, $d_A(z)$, which is the ratio of its physical size to the angle it subtends. It is sometimes called the "comoving angular diameter distance" and it is even sometimes written as $d_A(z)$. See [95] for a helpful summary of these often confusingly named quantities.

the bias (Equation 2.15). A 10^6 voxel data cube, for example, would take $\mathcal{O}(10^{18})$ computational steps to analyze. This is simply infeasible for next-generation radio interferometers and we have therefore endeavored to find a faster way to compute 21 power spectrum statistics.

2.3 Our Fast Method

To avoid the computational challenges of the LT method, we seek to exploit symmetries and simpler forms in certain bases of the various matrices out of which we construct our estimate of the 21 cm power spectrum and its attendant errors and window functions. In this section, we describe the mathematical and computational techniques we employ to create a fast and scalable algorithm.

Our fast method combines the following six separate ideas:

1. A Monte Carlo technique for computing the Fisher information matrix and the bias (Section 2.3.1).

2. An FFT-based technique for computing band powers using the $\mathbf{Q}^\alpha$ matrices (Section 2.3.2).

3. An application of the conjugate gradient method that eliminates the need to invert $\mathbf{C}$ (Section 2.3.3).

4. A Toeplitz matrix technique for multiplying vectors quickly by the constituent matrices of $\mathbf{C}$ (Section 2.3.4).

5. A combined FFT and spectral technique for preconditioning $\mathbf{C}$ to improve converge of the conjugate gradient method (Section 2.3.5)

6. A technique using spectral decomposition and Toeplitz matrices for rapid simulation of data cubes for our Monte Carlo (Section 2.3.6).

In this Section, we explain how all six are realized and how they fit into our fast method for power spectrum estimation. Finally, in Section 2.3.7, we verify the algorithm in an end-to-end test.

2.3.1 Monte Carlo Calculation of the Fisher Information Matrix

In order to turn the results of our quadratic estimator into estimates of the power spectrum with proper vertical and horizontal error bars, we need to be able to calculate the Fisher information matrix and the bias term. Instead of using the form of $\mathbf{F}$ in Equation 2.14 that the LT method employs, we take advantage of the relationship between $\mathbf{F}$ and $\widehat{\mathbf{q}}$ in Equation 2.21 that $\mathbf{F} = \langle \widehat{\mathbf{q}}\widehat{\mathbf{q}}^\mathsf{T} \rangle - \langle \widehat{\mathbf{q}} \rangle \langle \widehat{\mathbf{q}} \rangle^\mathsf{T}$. If we can generate a large number of simulated data sets $\mathbf{x}$ drawn from the same covariance $\mathbf{C}$ and then compute $\widehat{\mathbf{q}}$ from each one, then we can iteratively approximate $\mathbf{F}$ with a Monte Carlo. In other words, a solution to the problem of quickly calculating $\widehat{\mathbf{q}}$ also provides us with a way to estimate $\mathbf{F}$. What's more, the solution is trivially parallelizable; creating artificial data cubes and analyzing them can be done by many CPUs simultaneously.

In calculating $\mathbf{F}$, we can get b^α out essentially for free. If we take the average of all our $\widehat{\mathbf{q}}$ vectors, we expect to that

$$
\begin{aligned}
\langle \widehat{q}^\alpha \rangle &= \left\langle \frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^\mathsf{T} \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle) \right\rangle \\
&= \mathrm{tr}\left[\left\langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^\mathsf{T} \right\rangle \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \right] \\
&= \mathrm{tr}\left[\mathbf{Q}^\alpha \mathbf{C}^{-1} \right] = b^\alpha
\end{aligned}
\tag{2.34}
$$

in the limit where $\mathbf{S}$ is negligibly small. This implies that $\widehat{\mathbf{p}}$ can be written in an even simpler way:

$$
\widehat{p}^\alpha = \frac{1}{\sum_\gamma F^{\alpha\gamma}} \left(\widehat{q}^\alpha - \langle \widehat{q}^\alpha \rangle \right)
\tag{2.35}
$$

where, recall, $\mathbf{F}$ is calculated as the sample covariance of our $\widehat{\mathbf{q}}$ vectors. We therefore can calculate all the components of our power spectrum estimate and its error bars using a Monte Carlo.

In Section 2.3.7 we will return to assess how well the Monte Carlo technique works and its convergence properties. But first, we need to tackle the three impediments to computing $\widehat{q}^\alpha$ in Equation 2.10 quickly: generating a random $\mathbf{x}$ drawn from $\mathbf{C}$, computing $\mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle)$, and applying $\mathbf{Q}^\alpha$.

2.3.2 Fast Power Spectrum Estimation Without Noise or Foregrounds

If we make the definition that

$$\mathbf{y} \equiv \mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle) \tag{2.36}$$

to simplify Equation 2.10 to

$$q^\alpha = \mathbf{y}^\mathsf{T} \mathbf{Q}^\alpha \mathbf{y}, \tag{2.37}$$

we can see that even if we have managed to calculate $\mathbf{y}$ quickly, we still need to multiply it by a $N \times N$ element $\mathbf{Q}^\alpha$ matrix for each band power α. Though each $\mathbf{Q}^\alpha$ respects translation invariance that could make multiplying by vectors faster, there exists an even faster technique that can calculate every entry of $\widehat{\mathbf{p}}$ simultaneously using fast Fourier transforms.

To see that this is the case, we substitute Equation 2.9 into Equation 2.37, reversing the order of summation and integration and factoring the integrand:

$$\widehat{q}^\alpha = \frac{1}{2} \left[\int_{k_\parallel^\alpha - \Delta k_\parallel /2}^{k_\parallel^\alpha + \Delta k_\parallel /2} - \int_{-k_\parallel^\alpha + \Delta k_\parallel /2}^{-k_\parallel^\alpha - \Delta k_\parallel /2} \right] \int_{k_\perp^\alpha - \Delta k_\perp /2}^{k_\perp^\alpha + \Delta k_\perp /2}$$
$$\left(\sum_i y_i e^{i\mathbf{k}\cdot\mathbf{r}_i} \right) \left(\sum_j y_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \right) |\Phi(\mathbf{k})|^2 \frac{k_\perp \, d\theta dk_\perp \, dk_\parallel}{(2\pi)^3}. \tag{2.38}$$

The two sums inside the integral are very nearly discrete, 3D Fourier transforms. All that remains is to discretize the Fourier space conjugate variable $\mathbf{k}$ as we have already discretized the real space variable $\mathbf{r}$.

In order to evaluate the outer integrals, we approximate them as a sum over grid points in Fourier space. The most natural choice for discretization in $\mathbf{k}$ is one that follows naturally from the FFT of $\mathbf{y}$ in real space. If our box is of size $\ell_x \ell_y \ell_z$ and

broken into $n_x \times n_y \times n_z$ voxels[8] we have that

$$\mathbf{r}_j = \left(\frac{j_x \ell_x}{n_x}, \frac{j_y \ell_y}{n_y}, \frac{j_z \ell_z}{n_z} \right) \tag{2.39}$$

$$\text{where } j_x, j_y, j_z \in \left\{ -\frac{n_{x,y,z}}{2}, ..., 0, ..., \frac{n_{x,y,z}}{2} - 1 \right\}.$$

The natural 3D Fourier space discretization is

$$\mathbf{k}_m = \left(\frac{2\pi m_x}{\ell_x}, \frac{2\pi m_y}{\ell_y}, \frac{2\pi m_z}{\ell_z} \right) \tag{2.40}$$

$$\text{where } m_x, m_y, m_z \in \left\{ -\frac{n_{x,y,z}}{2}, ..., 0, ..., \frac{n_{x,y,z}}{2} - 1 \right\}$$

with a Fourier space voxel volume

$$(\Delta k)^3 = \frac{2\pi}{\ell_x} \times \frac{2\pi}{\ell_y} \times \frac{2\pi}{\ell_z}. \tag{2.41}$$

With this choice of discretization, we will simplify our integrals by sampling Fourier space with delta functions, applying the approximation in the integrand of Equation 2.38 that

$$1 \approx \sum_m \frac{(2\pi)^3 \delta^3(\mathbf{k} - \mathbf{k_m})}{\ell_x \ell_y \ell_z}. \tag{2.42}$$

This simplifies Equation 2.38 considerably:

$$\widehat{q}^\alpha = \frac{1}{2} \sum_m \left(\sum_i y_i e^{i\mathbf{k}_m \cdot \mathbf{r}_i} \right) \left(\sum_j y_j e^{-i\mathbf{k}_m \cdot \mathbf{r}_j} \right) \frac{\chi^\alpha(\mathbf{k}_m) |\Phi(\mathbf{k}_m)|^2}{\ell_x \ell_y \ell_z}. \tag{2.43}$$

If we define $\widetilde{y}_m \equiv \sum_j y_i e^{-i\mathbf{k}_m \cdot \mathbf{r}_j}$, then we can write $\widehat{\mathbf{q}}$ as:

$$\widehat{q}^\alpha \approx \frac{1}{2\ell_x \ell_y \ell_z} \sum_m \widetilde{y}_m^* \widetilde{y}_m \chi^\alpha(\mathbf{k}_m) |\Phi(\mathbf{k}_m)|^2$$

$$= \frac{1}{2\ell_x \ell_y \ell_z} \sum_m |\widetilde{y}_m|^2 \chi^\alpha(\mathbf{k}_m) |\Phi(\mathbf{k}_m)|^2. \tag{2.44}$$

This result makes a lot of sense: after all, the power spectrum is—very roughly

---

[8]For simplicity and consistency we assume that n_x, n_y, and n_z are all even and we take the origin the to be the second of the two center bins.

speaking—the data Fourier transformed, squared, and binned with an appropriate convolution kernel.

This is a very quick way to calculate $\widehat{\mathbf{q}}$ because we can compute $\widetilde{\mathbf{y}}$ in $\mathcal{O}(N \log N)$ time (if we already have $\mathbf{y}$) and then we simply need to add $|\widetilde{y}_m|^2$ for every m, weighted by the value of the analytic function $|\Phi(\mathbf{k}_m)|^2$ to the appropriate band power α.[9] Each value of $|\widetilde{y}_m|^2$ gets mapped uniquely to one value of α, so there are only N steps involved in performing the binning. Unlike in the LT method, we perform the calculation of $\widehat{q}^\alpha$ for all values of α simultaneously.

However, the FFT approximation to Q_{ij}^α from Equation 2.42 does not work very well at large values of $(\mathbf{r}_i - \mathbf{r}_j)$ because the discrete version of $\mathbf{Q}^\alpha$ does not sample the continuous version of $\mathbf{Q}^\alpha$ very finely. This can be improved by zero padding the input vector y_i, embedding it inside of a data cube of all zeros a factor of ζ^3 larger. For simplicity, we restrict ζ to integer values where $\zeta = 1$ represents no zero padding. By increasing our box size, we decrease the step size in Fourier space and thus the distance between each grid point in Fourier space where we sample k with delta functions. Repeating the derivation from Equations 2.39 through 2.44 yields:

$$\widehat{q}^\alpha \approx \frac{1}{2\ell_x \ell_y \ell_z \zeta^3} \sum_m |\widetilde{y}_m|^2 \chi^\alpha(\mathbf{k}_m) |\Phi(\mathbf{k}_m)|^2, \tag{2.45}$$

where $\widetilde{\mathbf{y}}$ has been zero padded and then Fourier transformed. This technique of power spectrum estimation scales as $\mathcal{O}(\zeta^3 N \log N)$, which is fine as long as ζ is small,[10] In Figure 2-3 we see how increasing ζ from 1 to 5 greatly improves accuracy.

---

[9]For simplicty, we choose band power spectrum bins with the same width as our Fourier space bins (before zero padding). This linear binning scheme makes plotting, which is typically logarithmic in the literature, more challenging. On the other hand, it better spreads out the number of Fourier space data cube bins assigned to each band power.

[10]Though not the computational bottleneck, this step is the most memory intensive; it involves writing down an array of $\zeta^3 N$ double-precision complex numbers. This can reach into the gigabytes for very large data cubes.
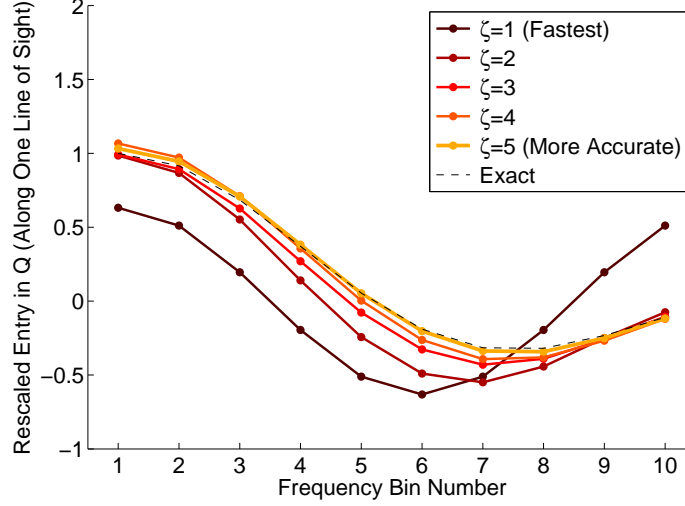
Figure 2-3: We use an FFT-based technique to approximate the action of the matrix $\mathbf{Q}^\alpha$ that encodes the Fourier transforming, binning, and pixelization factors. In this Figure, we show how the approximation improves with different factors of the zero padding parameter, ζ, while varying a single coordinate of one of the $\mathbf{Q}^\alpha$ matrices. For a fairly small value of ζ, the approximation is quite good, meaning that the binning and Fourier transforming step contributes subdomninantly to the complexity of the overall algorithm.

2.3.3 Inverse Variance Weighting with the Conjugate Gradient Method

We now know how to calculate $\widehat{q}^\alpha$ quickly provided that we can also calculate $\mathbf{y} \equiv \mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle)$ quickly. The latter turns out to be the most challenging part of the problem; we will address the various difficulties that it presents in this Section through Section 2.3.5. We take our inspiration for a solution from a similar problem that the WMAP team faced in making their maps. They employed the preconditioned conjugate gradient method to great success [164, 106].

The conjugate gradient method [90] is an iterative technique for solving a system of linear equations such as $\mathbf{Cy} = (\mathbf{x} - \langle \mathbf{x} \rangle)$. Although directly solving this system involves inverting the matrix $\mathbf{C}$, the conjugate gradient method can approximate the solution to arbitrary precision with only a limited number of multiplications of vectors by $\mathbf{C}$. If we can figure out a way to quickly multiply vectors by $\mathbf{C}$ by investigating the structure of its constituent matrices, then we can fairly quickly approximate $\mathbf{y}$. We

will not spell out the entire algorithm here but rather refer the reader to the helpful and comprehensive description of it in [199].

Whenever iterative algorithms are employed, it is important to understand how quickly they converge and what their rates of convergence depend upon. If we are trying to achieve an error ε on our approximation $\mathbf{y}_{\mathrm{CGM}}$ to $\mathbf{y}$ where

$$\varepsilon \equiv \frac{|\mathbf{C}\mathbf{y}_{\mathrm{CGM}} - (\mathbf{x} - \langle \mathbf{x} \rangle)|}{|\mathbf{x} - \langle \mathbf{x} \rangle|}. \tag{2.46}$$

and where $|\mathbf{x}| \equiv \left(\sum_i x_i^2\right)^{1/2}$ is the length of the vector $\mathbf{x}$, then the number of iterations required to converge (ignoring the accumulation of round-off error) is bounded by [199]:

$$n \leq \frac{1}{2}\sqrt{\kappa}\ln\left(\frac{2}{\varepsilon}\right) \tag{2.47}$$

where κ is the condition number of the matrix (not to be confused with κ used elsewhere as a spectral index), defined as the ratio of its largest eigenvalue to its smallest:

$$\kappa(\mathbf{C}) \equiv \frac{\lambda_{\mathrm{max}}(\mathbf{C})}{\lambda_{\mathrm{min}}(\mathbf{C})} \tag{2.48}$$

Because n only depends logarithmically on ε, the convergence of the conjugate gradient method is exponential. In order to make the algorithm converge in only a few iterations, it is necessary to ensure that κ is not too large. This turns out to be a major hurdle that we must overcome, because we will routinely need to deal with covariance matrices with $\kappa(\mathbf{C}) \approx 10^8$ or worse. This dynamic range problem is unavoidable; it comes directly from the ratio of the brightest foregrounds, typically hundreds of kelvin, to the noise and signal, typically tens of millikelvin. That factor, about 10^4, enters squared into the covariance matrices, yielding condition numbers of roughly 10^8. In Section 2.3.5 we will explain the efforts we undertake to mitigate this problem.

2.3.4 Foreground and Noise Covariance Matrices

Before we can go about ensuring that the conjugate gradient method converges quickly, we must understand the detailed structure of the constituent matrices of **C**. In particular, we will show that these matrices can all be multiplied by vectors in $\mathcal{O}(N \log N)$ time. We will first examine the new kind of foreground we want to include, resolved point sources, which will also provide a useful example for how the foreground covariances can be quickly multiplied by vectors.

2.3.4.1 Resolved Point Sources

Unlike LT, we do not assume that bright point sources have already been cleaned out of our map. Rather we wish to unify the framework for accounting for both resolved and unresolved foregrounds by inverse covariance weighting. This will allow us to directly calculate how our uncertainties about the fluxes and spectral indices of these point sources affect our ability to measure the 21 cm power spectrum.

In contrast to the unresolved point sources modeled by **U**, we model N_R bright resolved point sources as having known positions[11] with different fluxes S_n (at reference frequency ν_*) and spectral indices κ_n, neither of which is known perfectly. We assume that resolved point source contributions to **x** are uncorrelated with each other, so we can define an individual covariance matrix $\mathbf{R}_n$ for each point source. This means that our complete model for **R** is:

$$\mathbf{R} \equiv \sum_n \mathbf{R}_n. \tag{2.49}$$

Following LT, we can express the expected brightness temperature in a given voxel along the line of sight of the nth point source by a probability distribution for flux, $p_{S_n}(S')$, and spectral index, $p_{\kappa_n}(\kappa')$, that are both Gaussians with means S_n and κ_n and standard deviations σ_{S_n} and σ_{κ_n}, respectively. Following the derivation in LT,

---

[11] If the data cube is not overresolved, this assumption should be pretty good. If a point source appears to fall in two or more neighboring pixels, it could be modeled as two independent point sources in this framework. An even better choice would be to include the correlations between the two pixels, which would be quite strong. Modeling those correlations could only improve the results, since it would represent including additional information about the foregrounds, though it might slow down the method slightly. Not accounting for position uncertainty will cause the method to underestimate the "wedge" feature [50, 26, 230, 156, 225].

this yields:

$$
\begin{aligned}
\langle x_i \rangle_n &= \delta_{ii_n}(1.4 \times 10^{-3}\text{K})\left(\frac{\Omega_{pix}}{1\text{ sr}}\right)^{-1}\eta_i^{-2}\int_{-\infty}^{\infty}\left(\frac{S'}{1\text{ Jy}}\right)p_{S_n}(S')dS'\int_{-\infty}^{\infty}\eta_i^{-\kappa'}p_{\kappa_n}(\kappa')d\kappa' \\
&= \delta_{ii_n}(1.4 \times 10^{-3}\text{K})\left(\frac{S_n}{1\text{ Jy}}\right)\left(\frac{\Omega_{pix}}{1\text{ sr}}\right)^{-1}\times \eta_i^{-2-\kappa_n}\exp\left[\frac{\sigma_{\kappa_n}^2}{2}(\ln\eta_i)^2\right], \qquad (2.50)
\end{aligned}
$$

where again $\eta_i \equiv \nu_i/\nu_*$. Here δ_{ii_n} is a Kronecker delta that forces $\langle x_i \rangle$ to be zero anywhere other than the line of sight corresponding to the nth resolved point source. Likewise, we can write down the second moment:

$$
\begin{aligned}
\langle x_i x_j \rangle_n &= \delta_{ii_n}\delta_{jj_n}(1.4 \times 10^{-3}\text{K})^2(\eta_i\eta_j)^{-2}\times \\
&\quad \left(\int_{-\infty}^{\infty}\left(\frac{S'}{1\text{ Jy}}\right)^2 p_{S_n}(S')dS'\right)\left(\frac{\Omega_{pix}}{1\text{ sr}}\right)^{-2}\left(\int_{-\infty}^{\infty}(\eta_i\eta_j)^{-\kappa'}p_{\kappa_n}(\kappa')d\kappa'\right) \\
&= \delta_{ii_n}\delta_{jj_n}(1.4 \times 10^{-3}\text{K})^2(\eta_i\eta_j)^{-2-\kappa_n}\times \\
&\quad \left(\frac{S_n^2 + \sigma_{S_n}^2}{(1\text{ Jy})^2}\right)\left(\frac{\Omega_{pix}}{1\text{ sr}}\right)^{-2}\exp\left[\frac{\sigma_{\kappa_n}^2}{2}(\ln\eta_i\eta_j)^2\right] \qquad (2.51)
\end{aligned}
$$

where we assume $\sigma_{S_n} \approx 5\%$ of S_n and $\sigma_{\kappa_n} \approx 0.2$.

We know that $\langle x_i \rangle_n \langle x_j \rangle_n$ can be quickly multiplied by a vector because it is a rank 1 matrix. Therefore, in order to show that all of $\mathbf{R}$ can be quickly multiplied, we recast $\langle x_i x_j \rangle_n$ as the product of matrices that can be multiplied by a vector in $\mathcal{O}(N\log N)$ or faster. If we then ignore the constants and just look at the parts of this matrix that depend on coordinates, we have that:

$$
\begin{aligned}
\langle x_i x_j \rangle_n &\propto \delta_{ii_n}\delta_{jj_n}(\eta_i\eta_j)^{-2-\kappa_n}\exp\left[\frac{\sigma_{\kappa_n}^2}{2}(\ln\eta_i\eta_j)^2\right] \\
&= \delta_{ii_n}(\eta_i)^{-2-\kappa_n}\exp\left[\sigma_{\kappa_n}^2(\ln\eta_i)^2\right]\times \\
&\quad \exp\left[-\frac{\sigma_{\kappa_n}^2}{2}\left(\ln\frac{\eta_i}{\eta_j}\right)^2\right]\times \\
&\quad \delta_{jj_n}(\eta_j)^{-2-\kappa_n}\exp\left[\sigma_{\kappa_n}^2(\ln\eta_j)^2\right]. \qquad (2.52)
\end{aligned}
$$

This matrix can be separated into the product of three matrices: one diagonal matrix

that only depends on η_i, an inner matrix that includes the logarithm of a quotient of η_i and η_j in the exponent, and another diagonal matrix that only depends on η_j. The diagonal matrices can be multiplied by a vector in $\mathcal{O}(n_z)$. Moreover, because our cubes have redshift ranges $\Delta z < 0.5$ the frequencies at i and j are never very far apart, we can make the approximation that:

$$\ln\left(\frac{\eta_i}{\eta_j}\right) = \ln\left(\frac{\nu_i}{\nu_j}\right) \approx \frac{\nu_0 + \Delta\nu_i}{\nu_0 + \Delta\nu_j} - 1 = \frac{1}{\nu_0}(\Delta\nu_i - \Delta\nu_j) \qquad (2.53)$$

where $\Delta\nu_i \equiv \nu_i - \nu_0$ and ν_0 is a constant reference frequency close to both ν_i and ν_j. We choose the center frequency of the data cube to be ν_0. We can see now by combining Equations 2.52 and 2.53 that the inner matrix in our decomposition of the second moment depends only on the magnitude of the difference between ν_i and ν_j. In the approximation that the physical size of the data cube is small enough that frequencies map linearly to distances, this shows that $\mathbf{R}_n$ respects translational invariance along the line of sight.

Because the entries in this inner part of $\mathbf{R}_n$ only depend on differences in frequencies, the inner matrix is a diagonal-constant or "Toeplitz" matrix. Toeplitz matrices have the fortuitous property that they can be multiplied by vectors in $\mathcal{O}(N\log N)$, as we explain in Appendix 2.A. Therefore, we can multiply $\mathbf{R}_n$ by a vector in $\mathcal{O}(n_z \log n_z)$ and we can multiply $\mathbf{R}$ by a vector faster than $\mathcal{O}(N \log N)$.

We can understand this result intuitively as a consequence of the fact that the inner part of $\mathbf{R}_n$ is translationally invariant along the line of sight. Matrices that are translationally invariant in real space are diagonal in Fourier space. That we need to utilize this trick involving circulant and Toeplitz matrices is a consequence of the fact that our data cube is neither infinite nor periodic.

2.3.4.2 Unresolved Point Sources and the Galactic Synchrotron Radiation

Let us now take what we learned in Section 2.3.4.1 (and Appendix 2.A) to see if $\mathbf{U}$ can also be quickly multiplied by a vector. Looking back at Equation 2.28, we can see

that our job is already half finished; $(U_\perp)_{ij}$ only depends on the absolute differences between $(r_\perp)_i$ and $(r_\perp)_j$. Likewise, we can perform the exact same trick we employed in Equations 2.52 and 2.53 to write down the relevant parts of $(U_\parallel)_{ij}$ from Equation 2.30 with the approximation that $\Delta \nu_i$ is always small relative to ν_0:

$$(U_\parallel)_{ij} \propto \exp\left[-\frac{\sigma_{\kappa_n}^2}{2\nu_0^2}(\Delta\nu_i - \Delta\nu_j)^2\right]. \tag{2.54}$$

In fact, we can decompose $\mathbf{U}$ as a tensor product of three matrices sandwiched between two diagonal matrices:

$$\mathbf{U} = \mathbf{D_U}[\mathbf{U}_x \otimes \mathbf{U}_y \otimes \mathbf{U}_z]\mathbf{D_U}. \tag{2.55}$$

where all three inner matrices are Toeplitz matrices. When we wish to multiply $\mathbf{U}$ by a vector, we simply pick out one dimension at a time and multiply every segment of the data by the appropriate Toeplitz matrix (e.g. every line of sight for $\mathbf{U}_z$). All together, the three sets of multiplications can be done in $\mathcal{O}(n_x n_y n_f \log n_f) + \mathcal{O}(n_x n_f n_y \log n_y) + \mathcal{O}(n_y n_f n_x \log n_x) = \mathcal{O}(N \log N)$ time.

Moreover, since $\mathbf{G}$ has exactly the same form as $\mathbf{U}$, albeit with different parameters, $\mathbf{G}$ too can be multiplied by a vector in $\mathcal{O}(N \log N)$ time by making the same approximation that we made in Equation 2.53.

2.3.4.3 Instrumental Noise

Lastly, we return now to the form of $\mathbf{N}$ we introduced in Section 2.2.4.3. To derive a form we combine Equations 2.32 and 2.33. The details are presented in Appendix 2.B, so here we simply state the result:

$$\mathbf{N} = \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}} \mathbf{F}_\perp, \tag{2.56}$$

where $\mathbf{F}_\perp$ and $\mathbf{F}_\perp^\dagger$ are the unitary discrete 2D Fourier and inverse Fourier transforms and where:

$$\widetilde{N}_{lm} = \frac{\lambda^4 T_{\text{sys}}^2 j_0^2(k_{x,l}\Delta x/2) j_0^2(k_{y,l}\Delta y/2)}{A_{\text{ant}}^2 (\Omega_{\text{pix}})^2 n_x n_y \Delta\nu} \frac{\delta_{lm}}{t_l}. \tag{2.57}$$

Here, A_{ant} is the effective area of a single antenna, $\Delta\nu$ is the frequency channel width, l and m are indices that index over both uv-cells and frequencies, and t_l is the total observation time in a particular uv-cell at a particular frequency.

Because this matrix is diagonal, we have therefore shown that $\mathbf{N}$, along with $\mathbf{R}$, $\mathbf{U}$, and $\mathbf{G}$, can be multiplied by a vector in $\mathcal{O}(N \log N)$. We have summarized the results for all four matrices in Table 2.1.

2.3.4.4   Eliminating Unobserved Modes with the Psuedo-Inverse

In our expression for the noise covariance in Equation 2.57, we are faced with the possibility that t_l could be zero for some values of l, leading to infinite values of N_{ij}. Fourier modes with $t_l = 0$ correspond to parts of the uv-plane that are not observed by the instrument, i.e. to modes containing no cosmological information. We can completely remove these modes by means of the "psuedo-inverse" [213], which replaces $\mathbf{C}^{-1}$ in the expression $\mathbf{C}^{-1}(\mathbf{x} - \langle\mathbf{x}\rangle)$ and optimally weights all observed modes (this removal can itself be thought of as an optimal weighting—the optimal weight being zero). The psuedo-inverse involves $\mathbf{\Pi}$, a projection matrix ($\mathbf{\Pi}^{\dagger} = \mathbf{\Pi}$ and $\mathbf{\Pi}^2 = \mathbf{\Pi}$) whose eigenvalues are 0 for modes that we want to eliminate and 1 for all other modes. It can be shown [213] that the quantity we want to calculate for inverse variance weighting is not $\mathbf{C}^{-1}(\mathbf{x} - \langle\mathbf{x}\rangle)$ but rather the quantity where:

$$\mathbf{C}^{-1} \longrightarrow \mathbf{\Pi}\left[\mathbf{\Pi}\mathbf{C}\mathbf{\Pi} + \gamma(\mathbf{I} - \mathbf{\Pi})\right]^{-1}\mathbf{\Pi}. \tag{2.58}$$

In this equation, γ can actually be any number other than 0. The term in brackets in the above equation replaces the eigenvalues of the contaminated modes of $\mathbf{C}$ with γ. The outer $\mathbf{\Pi}$ matrices then project those modes out after inversion. In this paper, we take $\gamma = 1$ as the convenient choice for the preconditioner we will develop in Section 2.3.5.

The ability to remove unobserved modes is also essential for analyzing real data cubes produced by an interferometer. Interferometers usually produce so-called "dirty maps," which are corrected for the effects of the primary beam but have been con-

94

Covariance Matrix	Parallel to the Line of Sight	Perpendicular to the Line of Sight
R: Resolved Point Sources	Toeplitz symmetry	Diagonal in real space
U: Unresolved Point Sources	Toeplitz symmetry	Toeplitz symmetry
G: Galactic Synchrotron Radiation	Toeplitz symmetry	Toeplitz symmetry
N: Instrumental Noise	Diagonal in real space	Diagonal in Fourier space

Table 2.1: Due to the symmetries or approximate symmetries our models for the foreground and noise covariance matrices, they can all be multiplied by a vector in $\mathcal{O}(N \log N)$ time or faster. Summarized above are the reasons why each matrix can be quickly multiplied by. Either the matrices respect translation invariance and thus Toeplitz symmetry, their components are uncorrelated between lines of sight or frequencies, making them diagonal in real space, or they are uncorrelated and thus diagonal in Fourier space. In each case, the symmetries rely on the separabiltiy of the modeled covariance matrices into the tensor product of parts parallel or perpendicular to the line of sight.

volved by the synthesized beam, represented by the matrix $\mathbf{B}$:

$$\mathbf{x}_{\text{dirty map}} = \mathbf{B}\mathbf{x}. \tag{2.59}$$

To compute $\mathbf{x}$ for our quadratic estimator, we need to invert $\mathbf{B}$. Since the synthesized beam matrix is diagonal in Fourier space, this would be trivial were it not for unobserved baselines that make $\mathbf{B}$ uninvertable. This can be accomplished with the psuedoinverse as well, since the modes that would have been divided by 0 when inverting $\mathbf{B}$ are precisely the modes that we will project out via the psuedoinverse. We can therefore comfortably take

$$\mathbf{x} = \mathbf{F}_\perp^\dagger \mathbf{\Pi}[\mathbf{\Pi}\widetilde{\mathbf{B}}\mathbf{\Pi} + \gamma(\mathbf{I} - \mathbf{\Pi})]^{-1}\mathbf{\Pi}\mathbf{F}_\perp \mathbf{x}_{\text{dirty map}}, \tag{2.60}$$

where $\mathbf{B} \equiv \mathbf{F}_\perp^\dagger \widetilde{\mathbf{B}} \mathbf{F}_\perp$ and $\widetilde{\mathbf{B}}$ is diagonal.

The psuedo-inverse formalism can be usefully extended to any kind of mode we want to eliminate. One especially useful application would be to eliminate frequency channels contaminated by radio frequency interference or adversely affected by aliasing or other instrumental issues.

2.3.5 Preconditioning for Fast Conjugate Gradient Convergence

We have asserted that the quantity $\mathbf{y} \equiv \mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle)$ can be estimated quickly using the conjugate gradient method as long as the condition number $\kappa(\mathbf{C})$ is reasonably small. Unfortunately, this is never the case for any realistic data cube we might analyze. In Figure 2-4 we plot the eigenvalues of $\mathbf{C}$ and its constituent matrices for a small data cube (only $6 \times 6 \times 8$ voxels) taken from a larger, more representative volume. In this example, $\kappa(\mathbf{C}) \approx 10^8$, which would cause the conjugate gradient method to require tens of thousands of iterations to converge. This is typical; as we discussed in Section 2.3.3, values of around 10^8 are to be expected. We need to do better.
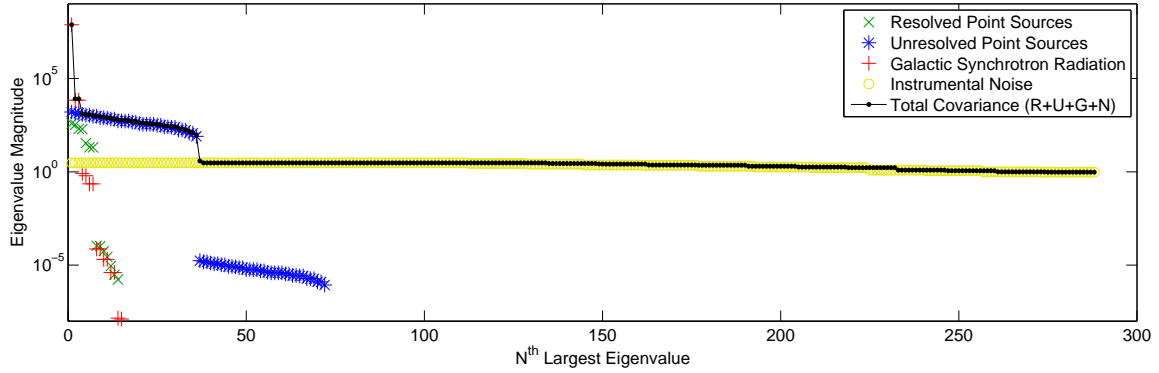
Figure 2-4: The distinct patterns in the eigenvalue spectrum of our covariance matrix provide an angle of attack for making the calculation of $\mathbf{C}^{-1}(\mathbf{x} - \langle\mathbf{x}\rangle)$ numerically feasible via preconditioning. The plotted eigenvalue spectra of the covariance for a very small data cube exemplifies many of the important characteristics of the constituent matrices. First, notice that the noise eigenvalue spectrum, while flatter than any of the others, is not perfectly flat. The condition number of N is related the ratio of the observing times in the most and least observed cell in the uv-plane. Sometimes this factor can be 10^3 or 10^4. Another important pattern to notice are the fundamental differences between the eigenvalue spectra of $\mathbf{U}$, $\mathbf{G}$, and $\mathbf{R}$. First off, $\mathbf{R}$ has mostly zero eigenvalues, because $\mathbf{R}$ is a block diagonal matrix with most of its blocks equal to zero. Second, despite the fact that $\mathbf{U}$ and $\mathbf{G}$ have nearly identical mathematical forms, $\mathbf{U}$ has stair-stepping eigenvalue spectrum while that of $\mathbf{G}$ is a much clearer exponential falloff. This is due to the much stronger correlations perpendicular to the line of sight in $\mathbf{G}$.

2.3.5.1 The Form of the Preconditioner

The core idea behind "preconditioning" is to avoid the large value of $\kappa(\mathbf{C})$ by introducing a pair of preconditioning matrices $\mathbf{P}$ and $\mathbf{P}^\dagger$. Instead of solving the linear system $\mathbf{C}\mathbf{y} = (\mathbf{x} - \langle \mathbf{x} \rangle)$, we solve the mathematically equivalent system:

$$\mathbf{C}'\mathbf{y}' = \mathbf{P}(\mathbf{x} - \langle \mathbf{x} \rangle), \tag{2.61}$$

where $\mathbf{C}' \equiv \mathbf{P}\mathbf{C}\mathbf{P}^\dagger$ and $\mathbf{y}' \equiv (\mathbf{P}^\dagger)^{-1}\mathbf{y}$. If we can compute $\mathbf{P}(\mathbf{x} - \langle \mathbf{x} \rangle)$ and, using the conjugate gradient method on $\mathbf{C}'$, we can solve for $\mathbf{y}'$ and thus finally find $\mathbf{y} = \mathbf{P}^\dagger\mathbf{y}'$. If $\mathbf{P}$ and $\mathbf{P}^\dagger$ are matrices that can be multiplied by quickly and if $\kappa(\mathbf{C}') \ll \kappa(\mathbf{C})$, then we can greatly speed up our computation of $\mathbf{y} = \mathbf{C}^{-1}(\mathbf{x} - \langle \mathbf{x} \rangle)$. Our goal is to build up preconditioning matrices specialized to the forms of the constituent matrices of $\mathbf{C}$. We construct preconditioners for $\mathbf{C} = \mathbf{N}$, generalize them to $\mathbf{C} = \mathbf{U} + \mathbf{N}$, and then finally incorporate $\mathbf{R}$ and $\mathbf{G}$ to build the full preconditioner.

The result is the following:

$$\mathbf{C}' = \mathbf{F}_\perp^\dagger \mathbf{P_U}\mathbf{P_\Gamma}\mathbf{P_N}(\mathbf{C})\mathbf{P_N^\dagger}\mathbf{P_\Gamma^\dagger}\mathbf{P_U^\dagger}\mathbf{F}_\perp. \tag{2.62}$$

Where $\mathbf{P_U}$, $\mathbf{P_\Gamma}$ and $\mathbf{P_N}$ and preconditioners for $\mathbf{U}$, $\mathbf{\Gamma} \equiv \mathbf{R} + \mathbf{G}$, and $\mathbf{N}$ respectively. A complete and pedagogical explanation of this preconditioner and the motivation for its construction and complex form can be found in Appendix 2.C. The definitions of the matrices can be found in Equations 2.93, 2.105, and 2.82 respectively.

Despite its complex form and construction, the procedure reduces $\kappa(\mathbf{C})$ by many orders of magnitude. In Figure 2-5, in explicit contrast to Figure 2-4, we see a demonstration of that effect.

2.3.5.2 Computational Complexity of the Preconditioner

In Appendix 2.C, we briefly discuss how the different steps in computing and applying this preconditioner scale with the problem size. If any of them scale too rapidly with
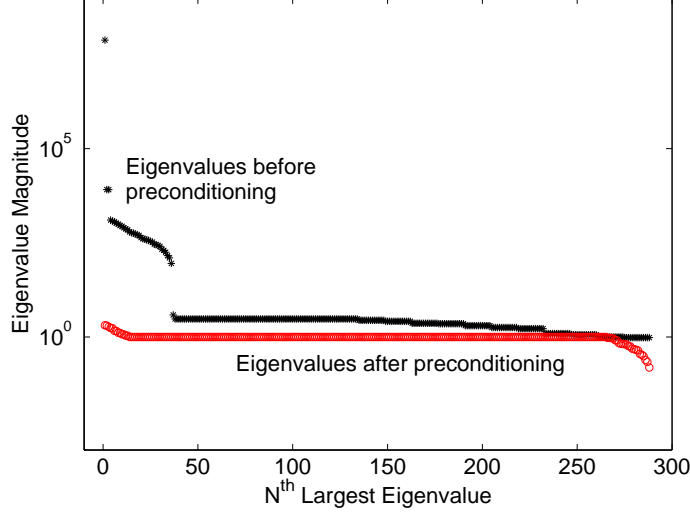
Figure 2-5: The preconditioner for the conjugate gradient method that we have devised significantly decreases the range of eigenvalues of **C**. Our preconditioner attempts to whiten the eigenvalue spectra of the constituent matrices of **C** sequentially, first **N**, then **R** and **G** together, and finally **U**. By preconditioning, the condition number $\kappa(\mathbf{C})$, the ratio of the largest to smallest eigenvalues, is reduced from over 10^8 to about 10^1.

N, we can quickly lose the computational advantage of our method over that of LT.[12]

First, let us enumerate the complexity of setting up the preconditioner for each matrix. $\mathbf{P_N}$ requires no setup since it only involves computing powers of the diagonal matrix $\widetilde{\mathbf{N}}$ (see Appendix 2.C.1). $\mathbf{P_U}$ requires the eigenvalue decomposition of $\mathbf{U}_z$, the component of **U** along the line of sight, which takes $\mathcal{O}(n_z^3)$ time (see Appendix 2.C.2).

We need the eigensystems of **R** and **G** to compute the eigensystem of $\mathbf{\Gamma}$ for $\mathbf{P_\Gamma}$ (see Appendix 2.C.3). **R** requires performing one eigenvalue decomposition of an $n_z \times n_z$ matrix for every resolved point source; that takes $\mathcal{O}(N_R n_z^3)$ time. **G** simply requires three eigenvalue decompositions: one for each matrix like those that appear for **U** in Equation 2.83 whose total outer product is **G**. Thus, the complexity is $\mathcal{O}(n_x^3) + \mathcal{O}(n_y^3) + \mathcal{O}(n_z^3)$.

Next, we need to compute the eigenvalues of $\mathbf{\Gamma}_{\perp,k}$, the components of $\mathbf{\Gamma}$ perpendicular to the line of sight corresponding to each of the "relevant" (i.e. much bigger

---

[12]This section may be difficult to follow without first reading Appendix 2.C. However, the key results can be found in Table 2.2.

Operation	Complexity
Compute $\mathbf{U}$ eigensystem	$\mathcal{O}(n_z^3)$
Compute $\mathbf{G}$ eigensystems	$\mathcal{O}(n_x^3) + \mathcal{O}(n_y^3) + \mathcal{O}(n_z^3)$
Compute $\mathbf{R}$ eigensystems	$\mathcal{O}(N_R n_z^3)$
Compute $\mathbf{\Gamma}$ eigensystems	$\mathcal{O}(m(\mathbf{G}_z)(n_x n_y)^3)$
Apply $\mathbf{P_N}$	$\mathcal{O}(N \log N)$
Apply $\mathbf{P_U}$	$\mathcal{O}(Nm(\mathbf{U}_z))$
Apply $\mathbf{P_\Gamma}$	$\mathcal{O}(Nm(\mathbf{G})) + \mathcal{O}(N N_R m(\mathbf{R}_n))$

Table 2.2: The computational complexity of setting up the preconditioner is, at worst, roughly $\mathcal{O}(N^2)$, though this operation only needs to be performed once. Even for large data cubes, this is not the rate-limiting step in power spectrum estimation. The computational complexity of applying the preconditioner ranges from $\mathcal{O}(N \log N)$ to $\mathcal{O}(NN_R)$. For large data cubes with hundreds of bright point sources, the preconditioning time is dominated by $\mathbf{P_\Gamma}$, which is in turn dominated by preconditioning associated with individual point sources. The computational complexity of the preconditioner therefore depends on the number of point sources considered "resolved," which scales with both field of view and with the flux cut. Here N_R is the number of resolved point sources in our field of view, n_d is the size of the box in voxels along the dth dimension, and m is the number of relevant eigenvalues of a matrix above the noise floor that need preconditioning.

than the noise floor) eigenvalues of $\mathbf{\Gamma}$ along the line of sight (see Appendix 2.C.3 for a more rigorous definition). Using the notation we develop in Appendix 2.C.2, we denote the number of relevant eigenvalues of a matrix $\mathbf{M}$ as $m(\mathbf{M})$. The number of times we need to decompose an $n_x n_y \times n_x n_y$ matrix is generally equal to the number of relevant eigenvalues of $\mathbf{G}_z$, since the number of relevant eigenvectors is almost always the same for $\mathbf{G}$ and $\mathbf{R}$. So we have then a computational complexity of $\mathcal{O}(m(\mathbf{G}_z)(n_x n_y)^3)$. Given the limited angular resolution of the experiment and the flat sky approximation, we generally expect n_x and n_y to be a good deal smaller than n_f, making this scaling more tolerable. All these scalings are summarized in Table 2.2.

Until now, all of our complexities have been $\mathcal{O}(N \log N)$ or smaller. Because these small incursions into bigger complexity classes are only part of the set-up cost, they are not intolerably slow as long as $m(\mathbf{G}_z)$ is small. This turns out to be true because the eigenvalue spectra of $\mathbf{R}_n$ and $\mathbf{G}_z$ fall off exponentially, meaning that we expect
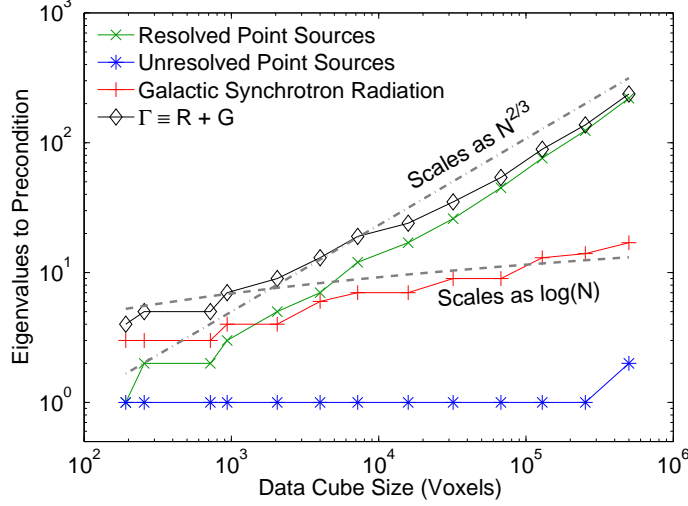
Figure 2-6: For large data cubes and a fixed definition of what constitutes a "bright" point source, the complexity of preconditioning is dominated by the number of resolved point sources. Specifically, the complexity of preconditioning for Γ scales as $N^{2/3}$ because the number of resolved point sources is simply proportional to the solid angle of sky surveyed, which scales with the survey volume (and thus number of voxels, assuming fixed angular and frequency resolution) to the $\frac{2}{3}$ power. This also confirms our assertion that the number of important eigenvalues of $\mathbf{G}$ and $\mathbf{U}$ should scale logarithmically with data cube size (albeit with a different prefactor). Each of the data cubes is taken from the same survey with the same ratio of width to depth. The number of eigenvalues to precondition is computed assuming an eigenvalue threshold of $\theta = 1$.

the number of relevant eigenvalues to grow only logarithmically. This is borne out in Figure 2-6 where we see exactly how the number of eigenvalues that need to be preconditioned scales with the problem size.

Let us now turn to a far more important scaling: that of multiplying the preconditioner by a vector. The set-up needs to be done only once per Fisher matrix calculation; the preconditioning needs to happen for every iteration of the conjugate gradient method. $\mathbf{P_N}$ is the easiest; we only ever need to perform a Fourier transform or multiply by a diagonal matrix. The complexity is merely $\mathcal{O}(N \log N)$. $\mathbf{P_U}$ only involves multiplying by vectors for each relevant eigenvalue of $\mathbf{U}_z$, so the total complexity is $\mathcal{O}(Nm(\mathbf{U}_z))$.

Finally, we need to assess the complexity of applying $\mathbf{P_\Gamma}$. When performing the eigenvalue decomposition of $\mathbf{\Gamma}_{\perp,k}$, we expect roughly the same number of eigenvalues

to be important that would have been important from $\mathbf{R}$ and $\mathbf{G}$ separately for that k index. Each of those eigenvectors takes $\mathcal{O}(N)$ time to multiply by a vector. So we expect to deal with $m(\mathbf{G})$ eigenvalues from $\mathbf{G}$ and one eigenvalue from each resolved point source for each relevant value of k, or about $N_R m(R_n)$. Applying $\mathbf{P_\Gamma}$ therefore is $\mathcal{O}(Nm(\mathbf{G})) + \mathcal{O}(NN_R m(\mathbf{R}_n))$. If we keep the same minimum flux for the definition of a resolved point source and if we scale our cube uniformly in all three spatial directions, then $N_R \propto N^{2/3}$.

This turns out to be the rate-limiting step in the entire algorithm. If we decide instead to only consider the brightest N_R to be resolved, regardless of box size, then applying $\mathbf{P_\Gamma}$ reduces to $\mathcal{O}(N \log N)$. Likewise, if we are only interested in expanding the frequency range of our data cube, the scaling also reduces to $\mathcal{O}(N \log N)$. We can comfortably say then that the inclusion of a model for resolved point sources introduces a complexity bounded by $\mathcal{O}(N \log N)$ and $\mathcal{O}(N^{5/3})$. We can see the precise computational effect of the preconditioner when we return in Section 2.4.3 to assess the overall scaling of the entire algorithm. These results are also summarized in Table 2.2.

2.3.5.3 Preconditioner Results

Choosing which eigenvalues are "relevant" in the constituent matrices of $\mathbf{C}$ and therefore need preconditioning depends on how these eigenvalues compare to the noise floor. In Appendix 2.C.2, we define a threshold θ which distinguishes relevant from irrelevant eigenvalues by comparing them to θ times the noise floor. Properly choosing a value for θ, the threshold below which we do not precondition eigenvalues of $\overline{\mathbf{U}}$ and $\overline{\mathbf{\Gamma}}$, presents a tradeoff. We expect that that too low of a value of θ will precondition inconsequential eigenvalues, thus increasing the conjugate gradient convergence time. We also expect that too large of a value of θ will leave some of the most important eigenvalues without any preconditioning, vastly increasing convergence time. Both of these expectations are borne out by our numerical experiments, which we present in Figure 2-7.

In this work, we choose $\theta = 1$ (all foreground eigenvalues above the noise floor are
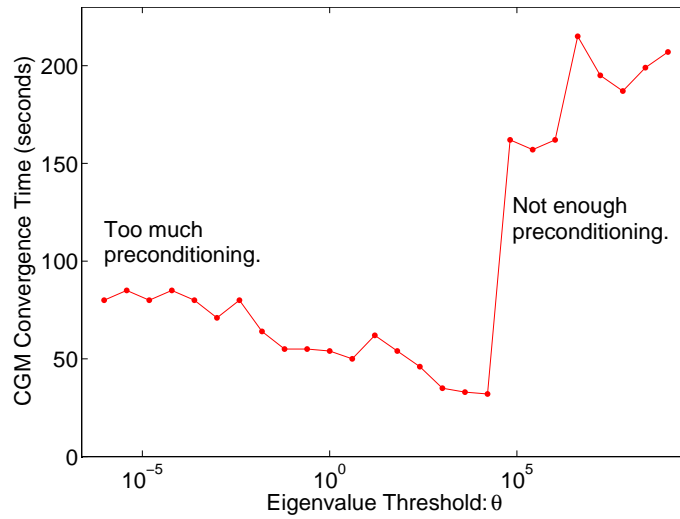
Figure 2-7: This plot shows how computational time scales with θ, the threshold for preconditioning, for the conjugate gradient method performed on an $N \approx 10^4$ voxel data cube. It appears that for this particular covariance matrix, a minimum exists near $\theta = 10^4$. At the minimum, the greater number of conjugate gradient iterations are balanced by quicker individual iterations (since each iteration involves less preconditioning). We can see from this plot that there exists a critical value of θ around 5×10^4 where the preconditioning of a small number of additional eigenvalues yields a large effect on the condition number of the resultant matrix. Without preconditioning, sufficiently large values of $\kappa(\mathbf{C})$ could also lead to the accumulation of roundoff error that prevents convergence of the conjugate gradient method.

preconditioned) for simplicity and to be sure that we are not skipping the preconditioning of any important foreground eigenvalues. One might also worry that more iterations of the algorithm provides more opportunity for round-off error to accumulate and prevent convergence, as has sometimes proven the case in our numerical experiments. For lengthy or repeated calculations of the Fisher matrix, it is wise to explore the performance of several levels of preconditioning, especially if it can garner us a another factor of 2 in speed.

2.3.6 Fast Simulation of Foregrounds and Noise

We concluded Section 2.3.1 with the fact that a Monte Carlo calculation of the Fisher matrix required the ability to compute $\widehat{\mathbf{q}}$ from many different realizations of the foregrounds and noise modeled by $\mathbf{C}$. In Sections 2.3.2 through 2.3.5, we have shown how to quickly calculate $\widehat{\mathbf{q}}$ from a data vector $\mathbf{x}$ using Equation 2.10.

But where does $\mathbf{x}$ come from? When we want to estimate the 21 cm temperature power spectrum of our universe, $\mathbf{x}$ will come from data cubes generated from real observations. But in order to calculate $\mathbf{F}$, which is essential both to measuring $\widehat{\mathbf{p}}$ and estimating the error on that measurement, we must first be able to create many realizations of $\mathbf{x}$ drawn from our models for noise and foregrounds that we presented in Section 2.3.4.

A mathematically simple way to draw $\mathbf{x}$ from the right covariance matrix is to create a vector $\mathbf{n}$ of independent and identically distributed random numbers drawn from a normal distribution with mean 0 and standard deviation 1. Then, it is easy to see that

$$\mathbf{x} \equiv \mathbf{C}^{1/2}\mathbf{n} \tag{2.63}$$

is a random vector with mean $\mathbf{0}$ and covariance $\mathbf{C}$. Unfortunately, computing $\mathbf{C}^{1/2}$ is just as computationally difficult as computing $\mathbf{C}^{-1}$.

In this last section of our presentation of our fast method for power spectrum estimation and statistics, we will explain how a vector can be created randomly with covariance $\mathbf{C}$. We do so by creating vectors randomly from each constituent matrix of

C, since each contribution to the measured signal is uncorrelated. In Section 2.3.6.5, we will demonstrate numerically that these simulations can be performed quickly while still being accurately described by the underlying statistics.

2.3.6.1 Resolved Point Sources

The simplest model to reproduce in a simulation is the one for resolved point sources, because the covariance was created from a supposed probability distribution over their true fluxes and spectral indices. We start with a list of point sources with positions and with a specified but uncertain fluxes and spectral indices. These fluxes can either come from a simulation, in which case we draw them from our source count distribution (Equation 2.29) and spectral indices from a Gaussian distribution, or from a real catalog of sources with its attendant error bars. The list of sources does not change over the course of calculating the Fisher matrix.

In either case, calculating a random $\mathbf{x_R}$ requires only picking two numbers, a flux and a spectral index, for each point source and then calculating a temperature in each voxel along that particular line of sight. The latter is easy, since we assume it is drawn from a Gaussian. The former can be quickly accomplished by numerically calculating the cumulative probability distribution from Equation 2.29 and inverting it. Each random $\mathbf{x_R}$ is therefore calculable in $\mathcal{O}(N_R n_z) < \mathcal{O}(N)$ time.

2.3.6.2 Unresolved Point Sources

We next focus on $\mathbf{U}$, which is more difficult. Our goal is to quickly produce a vector with specified mean and covariance. LT has already established what value we want for $\langle \mathbf{x_U} \rangle$ and $\langle \mathbf{x_G} \rangle$ with a calculation very similar to Equation 2.50. We need to figure out how to produce a vector with zero mean and the correct covariance.

One way around the problem of calculating $\mathbf{C}^{1/2}$ is to take advantage of the eigenvalue decomposition of the covariance matrix. That is because if $\mathbf{C} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$, where $\mathbf{Q}$ is the matrix that transforms into the eigenbasis and $\mathbf{\Lambda}$ is a diagonal matrix made up of the eigenvalues, then $\mathbf{C}^{1/2} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^\mathsf{T}$. We already found the few important eigenvalues of $\mathbf{U}$ for our preconditioner (see Section 2.C.2), so does this

technique solve our problem?

Yes and no. In the direction parallel to the line of sight, this technique works exceedingly well because only a small number of eigenvectors correspond to non-negligible eigenvalues. We can, to very good approximation, ignore all but the largest eigenvalues (which correspond to the first few "steps" in Figure 2-4.) We can therefore generate random unresolved point source lines of sight in $\mathcal{O}(n_z m(\mathbf{U}_z))$ with the right covariance.

A problem arises, however, when we want to generate $\mathbf{x_U}$ with the proper correlations perpendicular to the line of sight. Unlike the extremely strong correlations parallel to the line of sight, these correlations are quite weak. Weak correlations entail many comparable eigenvalues; in the limit that point sources were uncorrelated, $\mathbf{U}_x \otimes \mathbf{U}_y \to \mathbf{I}_\perp$ and all the eigenvalues would be 1 (though the eigenvectors would of course be much simpler too). Utilizing the same technique as above would require a total complexity of $\mathcal{O}(Nn_x n_y)$ time, which is slower than we would like.

However, the fact that both $\mathbf{U}_x$ and $\mathbf{U}_y$ are Toeplitz matrices allows us to use the same sort of trick we employed to multiply our Toeplitz matrices by vectors in Section 2.3.4.1 to draw random vectors from $\mathbf{U}_x \otimes \mathbf{U}_y$ [237]. It turns out that the circulant matrix in which we embed our covariance matrix must be positive-semidefinite for this technique to work. Although there exists such an embedding for any Gaussian covariance matrix, only Gaussians with coherence lengths small compared to the box size can be embedded in a reasonably small circulant matrix—exactly the situation we find ourselves in with $\mathbf{U}_\perp$. As such, we can generate random $\mathbf{x_U}$ vectors in $\mathcal{O}(Nm(\mathbf{U}_z) \log(n_x n_y)) \approx \mathcal{O}(N \log N)$.

2.3.6.3 Galactic Synchrotron Radiation

The matrix $\mathbf{G}$ differs from $\mathbf{U}$ primarily in the coherence length perpendicular to the line of sight. Unlike $\mathbf{U}$, $\mathbf{G}$ has only a small handful of important eigenvalues, which means that random $\mathbf{x_G}$ vectors can be generated in the same way we create line of sight components for $\mathbf{x_U}$ vectors, which we described above. Since $m(\mathbf{G})$ is so small (see Figure 2-4) and grows so slowly with data cube size (see Figure 2-6), we can

106

create random $\mathbf{x_G}$ vectors in approximately $\mathcal{O}(N)$.

2.3.6.4   Instrumental Noise

Finally, we turn to $\mathbf{N}$, which is also mathematically simple to simulate. First off, $\langle \mathbf{x_N} \rangle = 0$. Next, because $\mathbf{N}$ is diagonal in the Fourier basis, we can simply use Equation 2.63. Because $\mathbf{N} = \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}} \mathbf{F}_\perp$,

$$\mathbf{N}^{1/2} = \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}}^{1/2} \mathbf{F}_\perp, \tag{2.64}$$

which is computationally easy to multiply by $\mathbf{n}$ because $\widetilde{\mathbf{N}}$ is a diagonal matrix. The most computationally intensive step in creating random $\mathbf{x_N}$-vectors is the fast Fourier transform, which of course scales as $\mathcal{O}(N \log N)$.

2.3.6.5   Data Simulation Speed and Accuracy

Before we conclude this section and move on to the results of our method as a whole, we verify what we have claimed in the above sections: namely that we can quickly generate data cubes with the correct covariance properties. Figure 2-8 verifies the speed, showing that the algorithm is both fast and well-behaved for large data cubes.

In order to show that the sample covariance of a large number of random $\mathbf{x}$ vectors converges to the appropriate covariance matrix, we must first define a convergence statistic, ε. We are interested in how well the matrix converges relative to the total covariance matrix $\mathbf{C}$. For example, for $\mathbf{R}$ we choose:

$$\varepsilon(\widehat{\mathbf{R}}) \equiv \sqrt{\frac{\sum_{ij} \left| \widehat{R}_{ij} - R_{ij} \right|^2}{\sum_{ij} \left| C_{ij} \right|^2}} \tag{2.65}$$

where $\widehat{\mathbf{R}}$ is the sample covariance of n random $\mathbf{x_R}$ vectors drawn from $\mathbf{R}$. If each $\mathbf{x}$
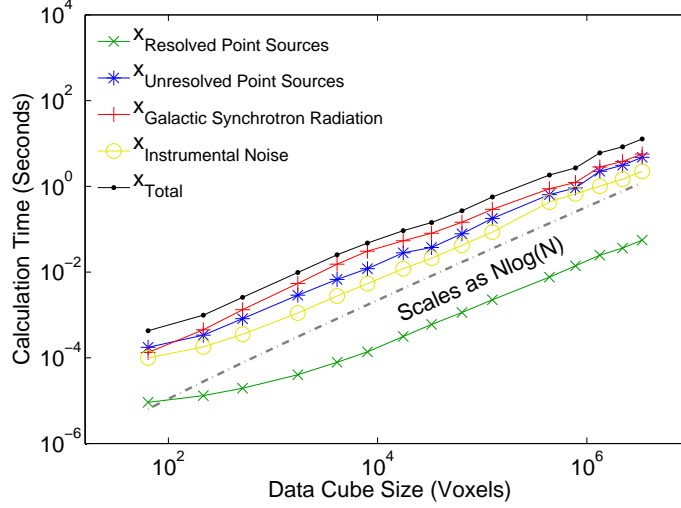
Figure 2-8: In order to estimate the Fisher matrix via a Monte Carlo, we need to draw random data cubes from our modeled covariance. Here we show that we can do so in $\mathcal{O}(N \log N)$ by plotting computational time as a function of problem size for generating a random $\mathbf{x}$ for each of the constituent sources of $\mathbf{x}$. In practice, generating random $\mathbf{x}$ vectors is never the rate-limiting step in calculating $\mathbf{F}$.

is a Gaussian random vector then the expected RMS value of ε is:

$$\sqrt{\left\langle \varepsilon(\widehat{\mathbf{R}})^2 \right\rangle} = \frac{1}{\sqrt{n}} \left[\frac{\sum_{i,j} R_{ij}^2 + (\mathrm{tr}\mathbf{R})^2}{\sum_{ij} C_{ij}^2} \right]. \tag{2.66}$$

In Figure 2-9, we see that all four constituent matrices of $\mathbf{C}$ converge like $n^{-1/2}$, as expected, with very nearly the prefactor predicted by Equation 2.66. We can be confident, therefore, in both the speed and accuracy of our technique for generating random vectors.

2.3.7 Method Accuracy and Convergence

Before we move on to discuss some of the results of our method, it is worthwhile to check that no unwarranted approximations prevent it from converging to the exact form of the Fisher information matrix in Equation 2.14. Since calculating $\mathbf{F}$ exactly can only be done in $\mathcal{O}(N^3)$ time, we perform this test in two parts.

First, we measure convergence to the exact Fisher matrix for a very small data cube with only $6 \times 6 \times 8$ voxels. Taking advantage of Equation 2.21, we generate an
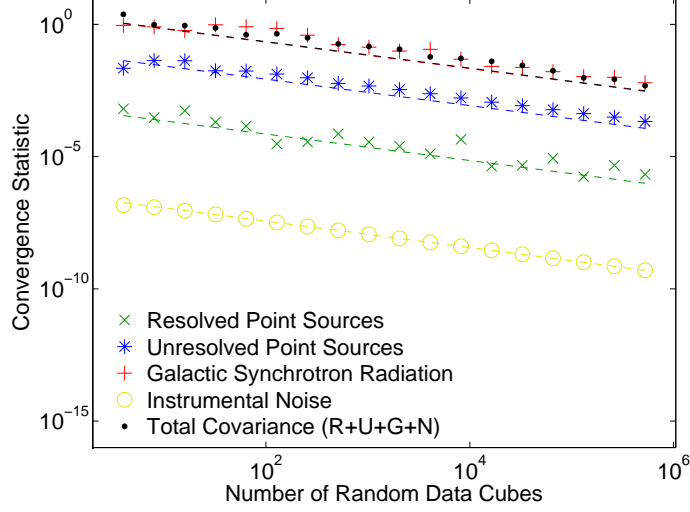
Figure 2-9: We verify that our technique for quickly generating random data cubes actually reproduces the correct statistics by generating a large number of such cubes and calculating their sample covariances. Plotted here is the error statistic detailed in Equation 2.65. The color-matched dotted lines are the expected convergences for correlated Gaussians from Equation 2.66.

estimate of $\mathbf{F}$, which we call $\widehat{\mathbf{F}}$, from the sample covariance of many independent $\widehat{\mathbf{q}}$ vectors. We compare these $\widehat{\mathbf{F}}$, which we calculate periodically along the course of the Monte Carlo, with the $\mathbf{F}$ that we calculated directly using Equation 2.14. As we show in Figure 2-10, the sample covariance of our $\widehat{\mathbf{q}}$ vectors clearly follows the expected $n^{-1/2}$ convergence to the correct result.

However, we are more concerned with the accuracy of the method for large data cubes which cannot be tackled by the LT method. Unfortunately, for such large data cubes, we cannot directly verify our result except in the case where $\mathbf{C} = \mathbf{I}$. In concert with other tests for agreement with LT, we also check that the method does indeed converge as $n^{-1/2}$ by comparing the convergence of subsets of the $\widehat{q}^{\alpha}$ vectors up to $n/2$ Monte Carlo iterations to the reference Fisher matrix, which we take to be the sample covariance of all n iterations. As we show in Figure 2-11, our expectation is borne out numerically.

109
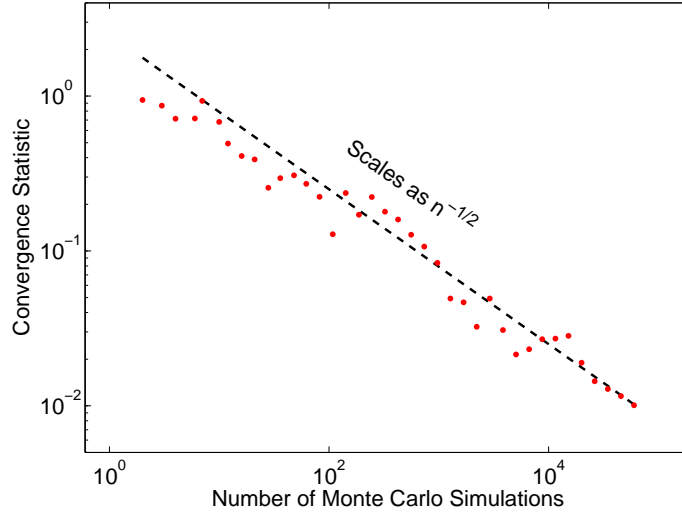
Figure 2-10: Our Monte Carlo converges to the correct Fisher matrix as $n^{-1/2}$, as expected. In this plot, we compare the sample covariance of many $\widehat{\mathbf{q}}$ vectors generated from small data cubes to an exact calculation of $\mathbf{F}$ by calculating the relative error of their diagonals.
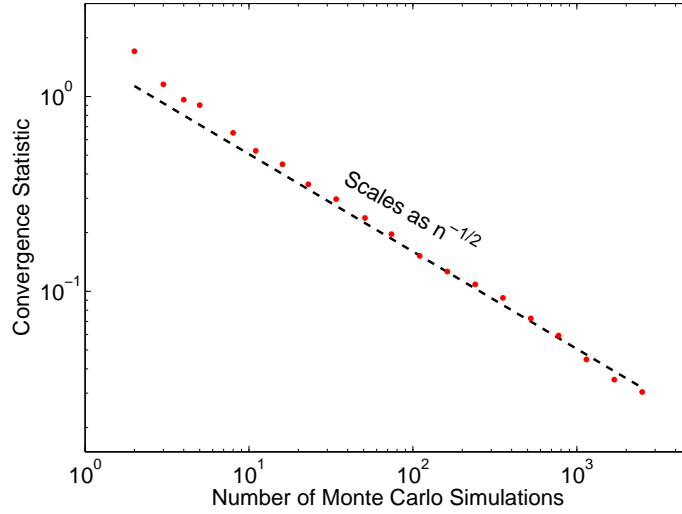
Figure 2-11: For large data cubes, the convergence of the sample covariance of our $\widehat{\mathbf{q}}$ vectors to $\mathbf{F}$ also nicely follows the expected $n^{-1/2}$ scaling. We perform this analysis on a data cube with 1.5×10^5 voxels analogously to that which we performed in Figure 2-10, except that we use the sample covariance of double the number of Monte Carlo iterations as our "true" Fisher matrix. This explains the artificially fast convergence we see in the last few points of the above plot.

2.3.8   Method Summary

We have constructed a technique that accelerates the LT technique to $\mathcal{O}(N \log N)$ and extends it to include bright point sources in, at worst, $\mathcal{O}(N^{5/3})$. We do so by generating random data vectors with the modeled foreground and noise covariances and calculating the Fisher information matrix via Monte Carlo. We are able to calculate individual inverse variance weighted power spectrum estimates quickly using the conjugate gradient method with a specially adapted preconditioner.

Our method makes a number of assumptions, most of which are not shared with the LT method. Our method can analyze larger data sets but at a slight loss of generality. Although we have mentioned these assumptions throughout this work, it is useful to summarize them in one place:

- Our method relies on a small enough data cube perpendicular to the line of sight that it can be approximated as rectilinear (see Figure 2-1).

- We approximate the natural log of the quotient of frequencies in the exponent of our point source covariance matrix by a leading-order Taylor expansion (Equation 2.53). This assumption makes the foreground covariances translationally invariant along the line of sight and thus amenable to fast multiplication using Toeplitz matrix techniques. This is a justified assumption as long as the coherence length of the foregrounds is much longer than the size of the box along the line of sight.

- Our ability to precondition our covariance matrix for the conjugate gradient method depends on the approximation that the correlation length of $\mathbf{U}$ perpendicular to the line of sight, due to weak spatial clustering of point sources, is not much bigger than the pixel size. For the purposes of preconditioning, we approximate $\mathbf{U}_{\perp}$ to be the identity (see Section 2.C.2). The longer the correlation length of $\mathbf{U}_{\perp}$, the longer the conjugate gradient algorithm will take to converge.

- Likewise, the speed of the preconditioned conjugate gradient algorithm depends

on the similarities of the eigenmodes of the covariances for $\mathbf{R}$, $\mathbf{U}$, and $\mathbf{G}$ along the line of sight. The more similar the eigenmodes are (though their accompanying eigenvalues can be quite different) the more the preconditioning algorithm can reduce the condition number of $\mathbf{C}$. We believe that this similarity is a fairly general property of models for foregrounds, though the introduction of a radically different foreground model might require a different preconditioning scheme.

- We assume that the number of Monte Carlo iterations needed to estimate the Fisher information matrix is not so large that that it precludes analyzing large data cubes. Because the process of generating more artificial $\widehat{\mathbf{q}}$ vectors is trivially parallelizable, we do not expect getting down to the requisite precision on the window functions to be an insurmountable barrier.

One common theme among these assumptions, especially the last three, is that the approximations we made to speed up the algorithm can be relaxed as long as we are willing to accept longer runtimes. This reflects the flexibility of the method, which can trade off speed for accuracy and vice versa.

2.4 Results

Now that we are confident that our method can accurately estimate the Fisher information matrix and can therefore calculate both power spectrum estimates from data and the attendant error bars and window functions, we turn to the first end-to-end results of the algorithm. In this Section, we demonstrate the power our method and the improvements that it offers over that of LT. First, in Section 2.4.1 we show that our technique reproduces the results of LT in the regions of Fourier space where they overlap. Then in Section 2.4.2, we highlight the improvements stemming from novel aspects of our method, especially the inclusion of the pixelization factor $|\Phi(\mathbf{k})|^2$ in $\mathbf{Q}^\alpha$ and $\mathbf{N}$ and the separation of point sources into resolved point sources ($\mathbf{R}$) and unresolved point sources ($\mathbf{U}$), by showing how different parts of our algorithm affect

F. In Section 2.4.3 we examine just how much faster our algorithm is than that of LT, and lastly, in Section 2.4.4 we forecast the cosmological constraining power of the 128-tile deployment of the MWA.

2.4.1 Comparison to Liu & Tegmark

First we want to verify that our method reproduces that of LT in the regions of Fourier space accessible to both methods. Figure 2-12 provides an explicit comparison to LT's Figures 2 and 8. These plots show the shaded regions representing their method and over-plotted, white-outlined contours representing ours. Both are on the same color scale. These plots show error bars in temperature units in $k_\perp$-$k_\parallel$ space and a selection of window functions in both the case where $\mathbf{C} = \mathbf{N}$ and $\mathbf{C} = \mathbf{U} + \mathbf{G} + \mathbf{N}$. They are generated from the same survey geometry with identical foreground and noise parameters. In the regions where the methods overlap, we see very good agreement between the two methods.

In addition to the modes shown in the shaded regions in Figure 2-12, the LT method can access Fourier modes longer than the box size, which we cannot. This is no great loss—these modes are poorly constrained by a single data cube. Moreover, they are generally those most contaminated by foregrounds; the low-$k_\perp$ modes will see heavy galactic synchrotron contamination while the low-$k_\parallel$ modes will be contaminated by types of the foregrounds. We imagine that very low-$k_\perp$ Fourier modes, those that depend on correlations between data cubes that cannot be joined without violating the flat-sky approximation, will still be analyzed by the LT method. Because our method can handle many more voxels, it excels in measuring both medium and high-k modes that require high spectral and spatial resolution.

2.4.2 Novel Effects on the Fisher Matrix

A simple way to understand the different effects that our forms of $\mathbf{C}$ and $\mathbf{Q}^\alpha$ have on the Fisher information matrix, especially the novel inclusions of $\mathbf{R}$ and $|\Phi(\mathbf{k})|^2$, is to build up the Fisher matrix component by component. In Figure 2-13 we do precisely
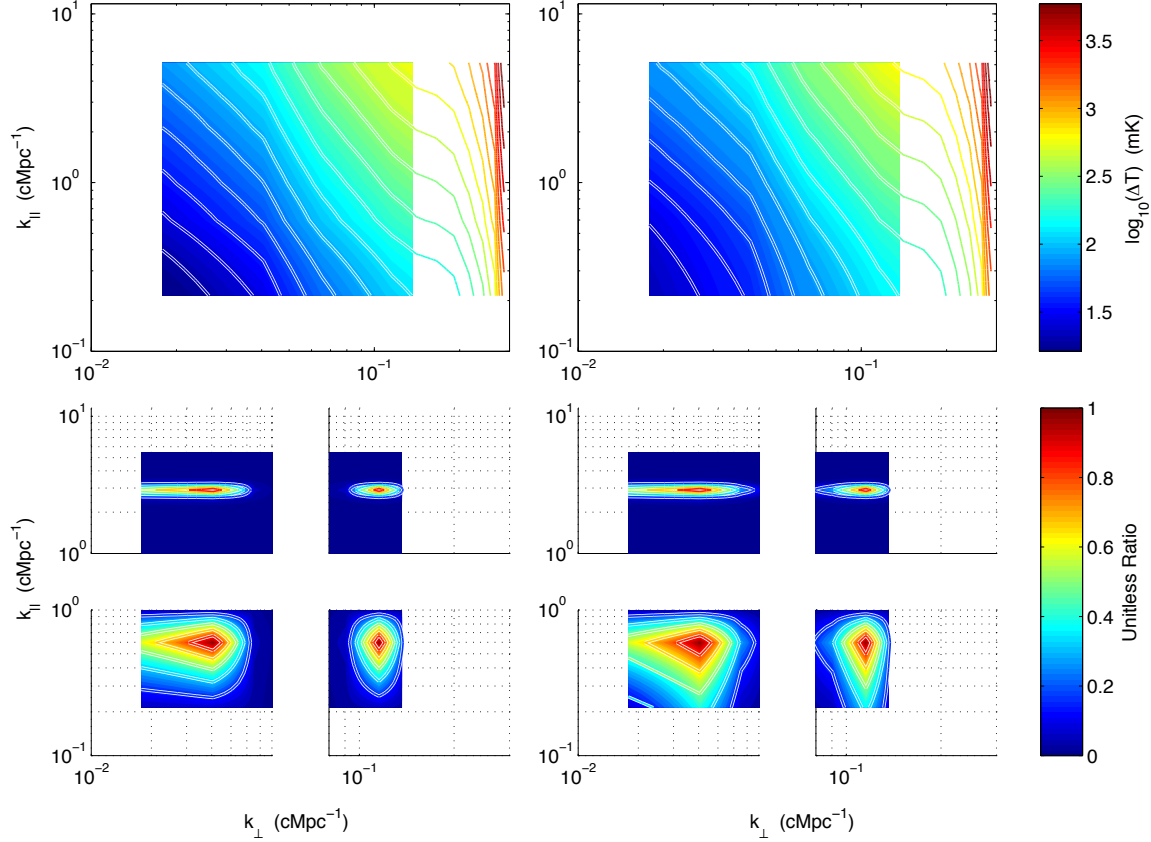
Figure 2-12: Our method faithfully reproduces the results of LT in the regions of Fourier space accessible to both methods. Here we recreate both the vertical error bar contours from LT's Figure 8 (top two panels) and a few selected window functions from LT's Figure 2 (bottom two groups of panels). The shaded regions represent the LT results; the white-outlined, colored contours are overplotted to show our results. Both are on the same color scale. Following LT, we have plotted both the case without foregrounds ($\mathbf{C} = \mathbf{N}$, left two panels) and the case with foregrounds ($\mathbf{C} = \mathbf{N} + \mathbf{U} + \mathbf{G}$, right two panels), which allows us to get a sense for the effects of foregrounds on our power spectrum estimates, error bars, and window functions.

that by plotting the diagonal elements of $\mathbf{F}$. These diagonal elements are related to the vertical error bars on our band powers. Large values of $F^{\alpha\alpha}$ correspond to band powers about which we have more information.

In the top two panels of Figure 2-13, we show the first novel effect that our method takes into account. In them, we can see how modeling the finite size of our voxels affects the information available in the case where $\mathbf{C} = \mathbf{I}$ (the color scale for these two panels only is arbitrary). In the top left panel, we have set $|\Phi(\mathbf{k})|^2 = 1$, which corresponds to the delta function pixelization of LT. We see that the amount of information depends only on $k_\perp$. This is purely a binning effect: our bins in $k_\perp$ are concentric circles with constant increments in radius; higher values of $k_\perp$ incorporate more volume in Fourier space, except at the high-$k_\perp$ edge where the circles are large enough to only include the corners of the data cube. In the top right panel, we see that including $|\Phi(\mathbf{k})|^2 \neq 1$ affects our ability to measure high-k modes, which depends increasingly on our real space resolution and is limited by the finite size of our voxels.

In the middle left panel, we now set $\mathbf{C} = \mathbf{N}$. In comparison to $\mathbf{C} = \mathbf{I}$, the new covariance matrix (and thus new vector $\mathbf{x}$ for the Monte Carlo calculation of $\mathbf{F}$), shifts the region of highest information to a much lower value of $k_\perp$. Though there are fewer Fourier modes that sample this region, there are far more baselines in the array configuration at the corresponding baseline length. Our noise covariance is calculated according to our derivation in Section 2.3.4.3 for 1000 hours of observation with the 128-tile deployment of the MWA [220].

We next expand to $\mathbf{C} = \mathbf{U} + \mathbf{N}$ for the middle right panel, where we have classified all point sources as "unresolved." In other words, we take S_{cut} in Equation 2.31 to be large (we choose 200 Jy, which is representative of some of the brightest sources at our frequencies). As we expect, smooth spectrum contamination reduces our ability to measure power spectrum modes with low values of $k_\parallel$. This is because of the exponentially decaying eigenvalue spectrum of $\mathbf{U}_\parallel$, most of which is smaller than the eigenvalues of $\mathbf{N}$. The effect is seen across $k_\perp$ because the characteristic clustering scale of unresolved point sources is smaller than the pixel size; localized structure in real space corresponds to unlocalized power in Fourier space.
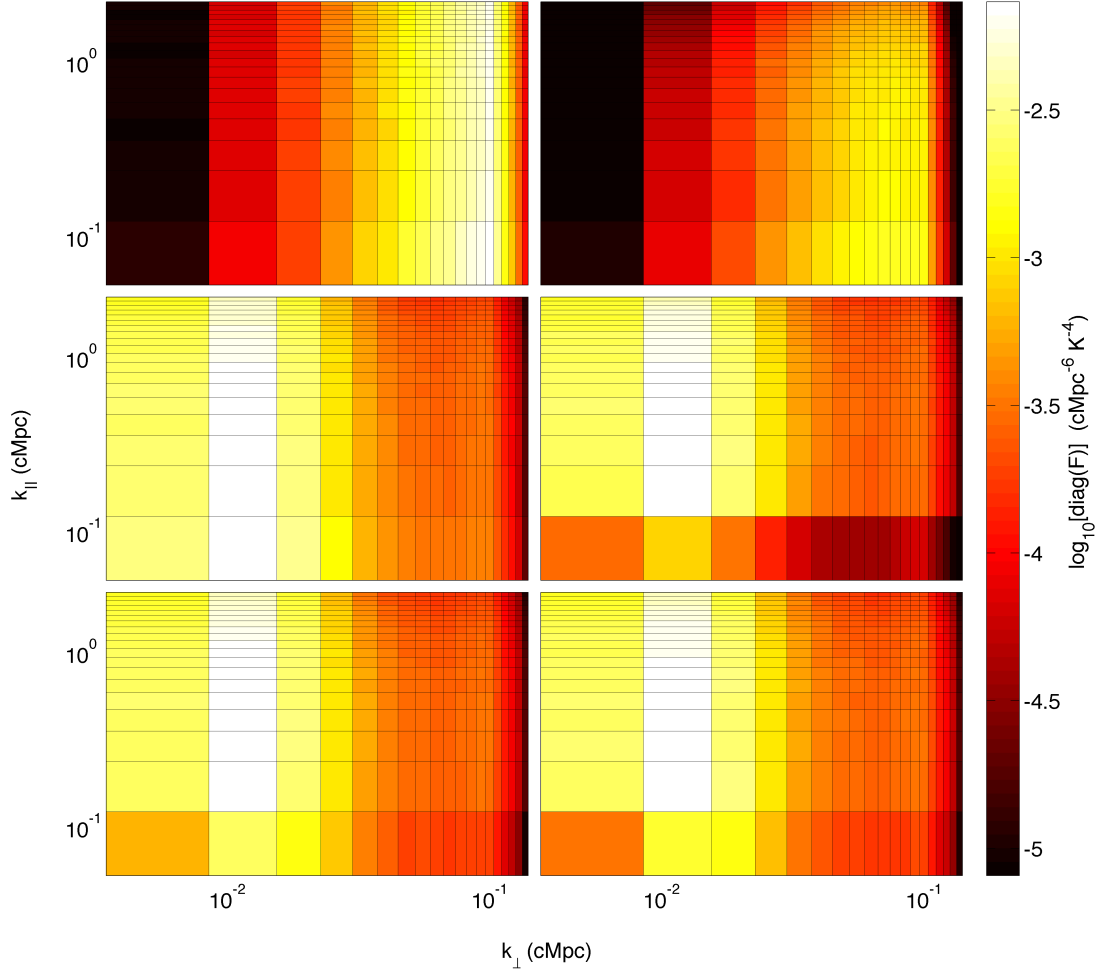
Figure 2-13: The Fisher information matrix provides a useful window into understanding the challenges presented to measuring the $21\,\mathrm{cm}$ power spectrum by the various contaminants. In this figure, we add these effects one by one, reading from left to right and top to bottom, to see how the diagonal of the Fisher matrix (expanded along the $k_\perp$ and $k_\parallel$ directions), is affected. Brighter regions represent, roughly speaking, more information and thus smaller error bars (in power spectrum units). We comment in more detail on each panel individually in Section 2.4.2, including upon the advantages of the novel aspects of our technique. *Top left:* the covariance matrix is taken to be the identity and pixelization effects on $\mathbf{Q}^\alpha$ are ignored. *Top right:* the pixelization factor $|\Phi(\mathbf{k})|^2$ is included and not set to 1. *Middle left:* the noise expected from 1000 hours of observation with the MWA 128-tile configuration is included. *Middle right:* all point sources (up to 200 Jy) are modeled as unresolved; all information about their positions is ignored. *Bottom left:* resolved point sources are included in the model, with all point sources dimmer than 100 mJy considered unresolved. *Bottom right:* in addition to bright point sources, galactic synchrotron is also included.

116

In the bottom left panel, we have included information about the positions of roughly 200 resolved point sources above 100 mJy, with random fluxes drawn from our source count distribution (Equation 2.29) and random spectral indices drawn from a Gaussian centered on $\kappa_n = 0.5$ with a width of 0.15. By doing this, we reduce S_{cut} in our model for $\mathbf{U}$ down to 100 mJy. Including all this extra information—positions, fluxes, flux uncertainties, spectral indices, and spectral index uncertainties—provides us with significantly more Fisher information at low-$k_\parallel$ where foregrounds dominate and thus smaller errors on those modes. Additionally, by incorporating resolved point sources as part of our inverse covariance weighting, we no longer have to worry about forward propagating errors from any external point-source subtraction scheme. In the left panel of Figure 2-14 we see the ratio of this panel to the middle right panel.

Finally, in the bottom right panel of Figure 2-13 we show the effect of including Galactic synchrotron radiation. Adding $\mathbf{G}$ has the expected effect; we already know that $\mathbf{G}$ has only a few important eigenmodes which correspond roughly to the lowest Fourier modes both parallel and perpendicular to the line of sight. As a result, we only see a noticeable effect in the bottom left corner of the $k_\perp$-$k_\parallel$ plane; we include the ratio of the two figures in the right panel of Figure 2-14 for clarity. Otherwise, our Galaxy has very little effect on the regions of interest. In fact, the similarity between the this panel and the middle left panel tells us something very striking: in the regions of Fourier space that our data most readily probes, foregrounds (once properly downweighted) should not prove an insurmountable obstacle to power spectrum estimation.

The set of plots in Figure 2-13 is useful for developing a heuristic understanding of how noise and foregrounds affect the regions in which we can most accurately estimate the 21 cm power spectrum. With it, we can more easily identify the precise regions of k-space that we expect to be minimally contaminated by foregrounds and can thus tailor our instruments and our observations to the task of measuring the 21 cm power spectrum.
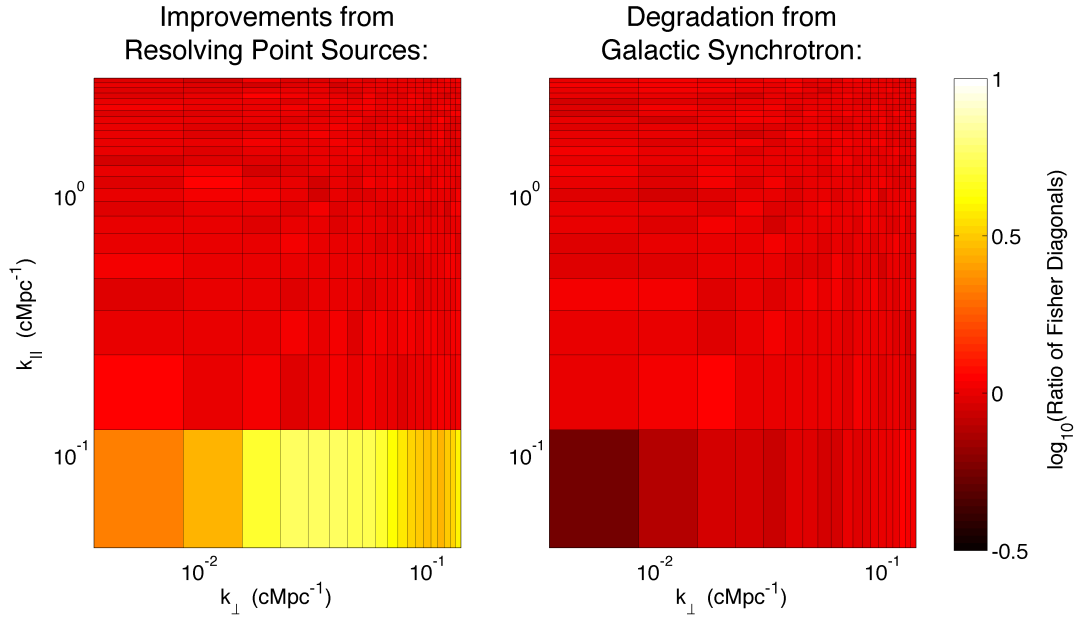
Figure 2-14: By comparing the Fisher matrices arising from the various covariance models we explore in Section 2.4.2 and Figure 2-13, the precise improvements are brought into sharper focus. In the left panel, we can see the information we gain by explicity including resolved point sources. Shown here is the ratio of the bottom left panel of Figure 2-13 to the middle right panel. By taking into account precise position, flux, and spectral information and uncertainties, we improve our ability to measure the power spectrum at the longest scales parallel to the line of sight, effectively ameliorating the effects of foregrounds. In the right panel, we see the remarkably small effect that the galactic synchrotron radiation has on our abilty the measure the 21 cm power spectrum. Shown here is the ratio of the bottom right panel of Figure 2-13 to the botton left. Because we take spatial information into account, the strong spatial and spectral coherence of the signal from our Galaxy is confined to the bottom left corner of the $k_\parallel$-$k_\perp$ plane.
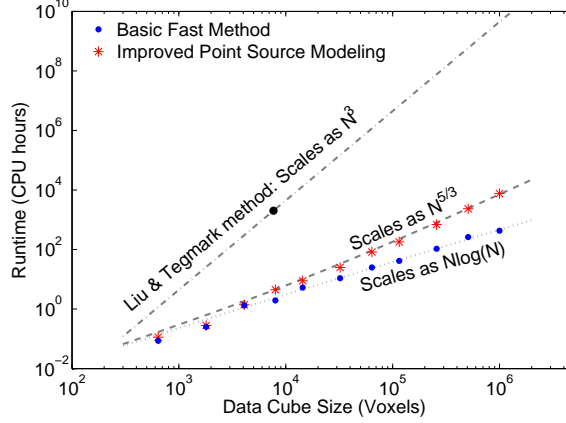
Figure 2-15: Our algorithm scales with the number of voxels, N, as $\mathcal{O}(N \log N)$ in the best case and as $\mathcal{O}(N^{5/3})$ in the worst, depending on the treatment of bright point sources. If we choose to ignore all information about the location, brightness, and spectra of bright point sources, we can estimate the power spectrum in $\mathcal{O}(N \log N)$. If we choose to take into account this extra information, the algorithmic complexity increases to $\mathcal{O}(N N_R)$, where N_R is the number of bright, resolved point sources. For a fixed minimum flux for "bright" sources, this leads to $\mathcal{O}(N^{5/3})$ complexity for uniform scaling in all three dimensions. Both scenarios represent a major improvement over the LT method, which scales as $\mathcal{O}(N^3)$.

2.4.3 Computational Scaling of the Method

Now that we understand how our technique works, we want to also see that it works as quickly as promised by and achieves the desired computational speed up over the LT method. Specifically, we want to show that we can achieve the theoretical $\mathcal{O}(N \log N)$ performance in reproducing the results of LT. We also want to better understand the computational cost of including resolved point sources so as to compare that cost to benefits outlined in Section 2.4.2. We have therefore tested the algorithm's speed for a wide range of data cube sizes; we present the results of that study in Figure 2-15.

In this figure, we show the combined setup and runtime for power spectrum estimates including 1000 Monte Carlo simulations of $\widehat{\mathbf{q}}$ for estimating the Fisher matrix on a single modern CPU. For each successive trial, we scale the box by the same ratio in all three dimensions. Because we maintain a fixed flux cut, increasing the linear size of the box by a factor of two increases the number of resolved point sources in the box by a factor of 4 and the number of voxels by a factor of 8. With any more than a few point sources, the computational cost becomes dominated by point sources,

leading to an overall complexity of $\mathcal{O}(N N_R)$. In this case, the largest data cubes include about 400 point sources over a field of about 50 square degrees, accounting for about 15% of the lines of sight in the data cube.

For any given analysis project, these exists a trade-off between including additional astrophysical information into the analysis and the computational complexity of that analysis; at some point the marginal cost of a slower algorithm exceeds marginal benefit of including more bright point sources. It is beyond the scope of this paper to prescribe a precise rubric for where to draw the line between resolved and unresolved point sources. However, we can confidently say that the algorithm runs no slower than $\mathcal{O}(N^{5/3})$ and can often run at or near $\mathcal{O}(N \log N)$ if only the brightest few point sources are treated individually.

2.4.4 Implications for Future Surveys

Though the primary purpose of this paper is to describe an efficient method for 21 cm power spectrum analysis, our technique enables us immediately to make predictions about the potential performance of upcoming instruments. In this section we put all our new machinery to work in order to see just how well the upcoming 128-tile deployment of the MWA can perform.

We envision 1000 hours of integration on a field that is 9° on each side, centered on $z = 8$ with $\Delta z = 0.5$. With a frequency resolution of 40 kHz and an angular resolution of 8 arcminutes, our data cube contains over 10^6 voxels. We completed over 1000 Monte Carlo iterations on our 12 core server in about one week. We use the foreground parameters outlined above in Sections 2.2.4 and 2.3.4. In Figures 2-16 through 2-18, we show the diagonal elements of the Fisher matrix we have calculated, the temperature power spectrum error bars, and a sampling of window functions.

In Figure 2-16 we plot the diagonal elements of the Fisher matrix, which are related directly to the power spectrum errors. Drawing from the discussion in Section 2.4.2, we can see clearly the effects of the array layout (and thus noise), of foregrounds (included resolved point sources), and of pixelization. Interestingly, until pixelization effects set in at the highest values of $k_\parallel$, the least contaminated region spans a large
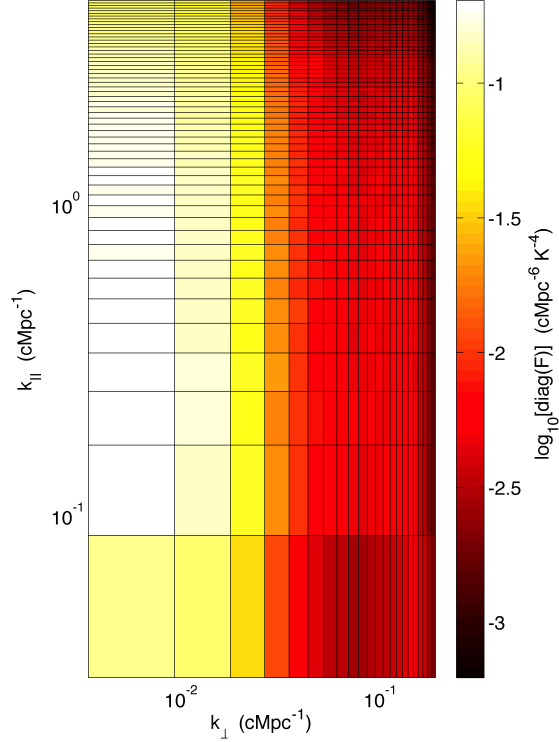
Figure 2-16: The diagonal of the Fisher matrix predicted for 1000 hours of observation with the MWA with 128 tiles shows the region of power spectrum space least contaminated by noise and foregrounds. Noise, and thus array layout, dominates the shape of the region of maximum information, creating a large, vertical region at a value of $k_\perp$ corresponding to the typical separation between antennas in the compact core of the array. The contaminating effects of the foregrounds are clearly visible at low-$k_\parallel$.
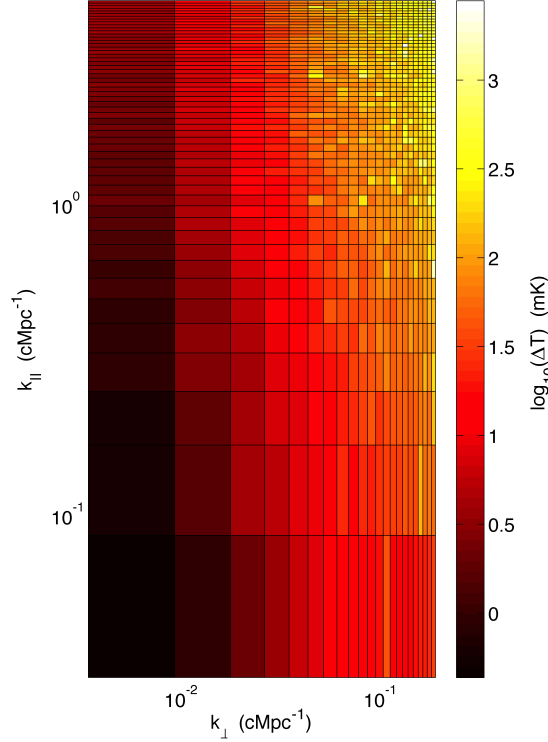
Figure 2-17: The expected error bars in temperature units on decorrelated estimates of the power spectrum highlight a sizable region of k-space where we expect to be able to use the MWA with 128 tiles to detect a fiducial 10 mK signal with a signal to noise ratio greater than 1. Perhaps surprisingly, the smallest error bars are still on the smallest k modes acessible by our method, though some of them are contaminated by large foregrounds. This is because our conversion to temperature units includes a factor of $(k_\perp^2 k_\parallel)^{1/2}$, which accounts for the difference between this Figure and Figure 2-16. From the shape of the region of smallest error, we can better appreciate the extent to which noise and our array layout determines where in k-space we might expect to be able to detect the EoR. The noisiness at high-k is due to Monte Carlo noise and can be improved with more CPU hours.

range of values of $k_\parallel$. One way of probing more cosmological modes is to increase the frequency resolution of the instrument. The number of modes accessible to the observation scales with $(\Delta\nu)^{-1}$, though the amplitude of the noise scales scales with $(\Delta\nu)^{-1/2}$. As long as the noise level is manageable and the cosmological signal is not dropping off too quickly with k, increasing the frequency resolution seems like a good deal.

In Figure 2-17 we show the vertical error bars that we expect on power spectrum estimates in temperature units. The most important fact about this plot is that there

is a large region where we expect that vertical error bars will be sufficiently small that we should be able to detect a 10 mK signal with signal to noise greater than 1. This is especially the case at fairly small values of k, which is surprising since these k modes were supposed to be the most contaminated by foregrounds. There are two reasons why this happens.

First, the conversion to temperature units (Equation 2.25) introduces a factor of $(k_\perp^2 k_\parallel)^{1/2}$ that raises the error bars for larger values of k. Second, the strongest foreground modes overlap significantly with one of the $k = 0$ modes of the discrete Fourier transform, which we exclude for our power spectrum estimate (this is just the average value of the cube in all three directions, which is irrelevant to an interferometer that is insensitive to the average).

Another way to think about it is this: because the coherence length of the foregrounds along the line of sight is much longer than the size of any box small enough to comfortably ignore cosmological evolution, we expect that the most contaminated Fourier mode will be precisely the one we ignore. Unlike the LT method, our method cannot easily measure modes much longer than the size of the data cube. Along the line of sight, these modes have very wide window functions and are the most contaminated by foregrounds. Perpendicular to the line of sight, these modes are better measured by considering much larger maps where the flat sky approximation no longer holds. For the purposes of measuring these low-k modes, the LT method can provide a useful complement to ours. Large-scale modes from down-sampled maps can be measured by LT; smaller-scale modes from full-resolution maps can be measured by our method. Then both can be combined to estimate the power spectrum across a large range of scales.

And finally, in Figure 2-18 we show many different window functions for a selection of values of $k_\perp$ and $k_\parallel$ that spans the whole space. In general, these window functions are quite narrow, meaning that each band power measurement probes only a narrow range of scales. The widest windows we see look wide for two reasons. First, linearly separated bins appear wider at low-k when plotted logarithmically. Second, foregrounds cause contaminated bins to leak into nearby bins, especially at low-$k_\parallel$ and
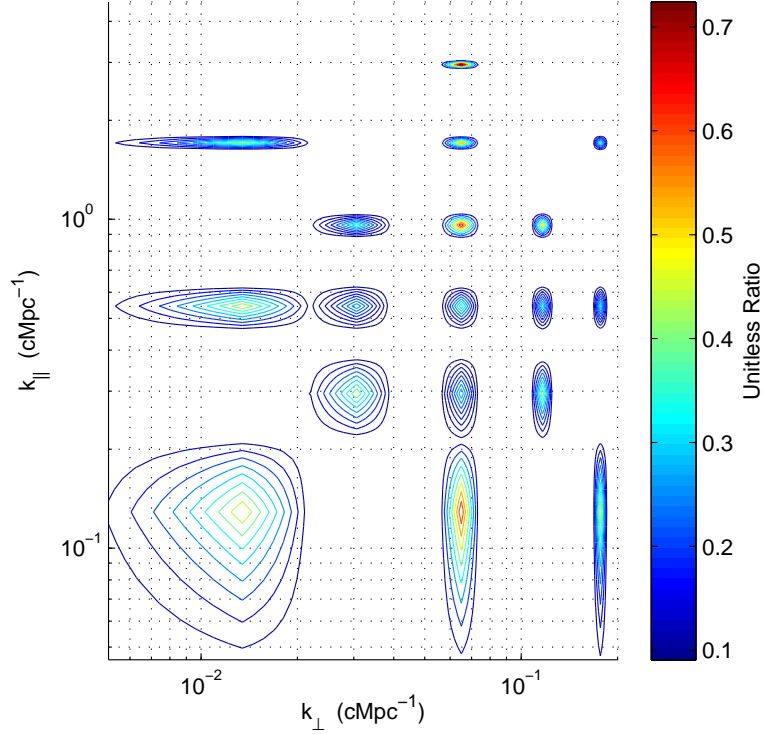
Figure 2-18: We can see from a sampling of window functions that our band power spectrum estimates $\widehat{p}^\alpha$ represent the weighted averages of p^α over a narrow range of scales, especially at higher values of $k_\perp$ and $k_\parallel$. The widest window functions can be attributed to binning (with linearly binned data, low-k bins look larger on logarithmic axes) and to foregrounds. This is good news, because it will enable us to accurately make many independent measurements of the power spectrum and therefore better constrain astrophysical and cosmological parameters.

moderate $k_\perp$. We saw hints of this effect in Figure 2-12 when comparing noise-only simulations to simulations with both noise and foregrounds.

In the vast majority of the $k_\perp$-$k_\parallel$ plane, the window functions seem to be dominated by the central bin and neighbors. Except for edge cases, no window function has contributions exceeding 10% from bins outside the central bin and its nearest neighbors. This means that we should be happy with our choice of Fourier space binning, which was designed to have bin widths equal to those of our data cube before zero padding. We also know that significantly finer binning would be inappropriate, so we do not have to worry about the tradeoff between fine binning of the power spectrum and the inversion of the Fisher matrix. Therefore, with the 128-tile deployment of the MWA, we can be confident that our estimates of the power spectrum correspond to distinct modes of the true underlying power spectrum.

2.5 Conclusions

With this paper, we have presented an optimal algorithm for 21 cm power spectrum estimation which is dramatically faster than the Liu & Tegmark (LT) method [120], scaling as $\mathcal{O}(N \log N)$ instead of $\mathcal{O}(N^3)$, where N is the number of voxels in the 3D sky map. By using the inverse variance weighted quadratic estimator formalism adapted to 21 cm tomography by the LT method, we preserve all accessible cosmological information in our measurement to produce the smallest possible error bars and narrow window functions. Moreover, our method can incorporate additional information about the brightest point sources and thus further reduce our error bars at the cost of some—but by no means all—of that computational advantage. Our method is highly parallelizable and has only modest memory requirements; it never needs to store an entire $N \times N$ matrix.

Our method achieves this computational speed-up for measuring power spectra, error bars, and window functions by eliminating the time-consuming matrix operations of the LT method. We accomplish this using a combination of Fourier, spectral, and Monte Carlo techniques which exploit symmetries and other physical properties

of our models for the noise and foregrounds.

We have demonstrated the successful simulation of error bars and window functions for the sort of massive data set we expect from the upcoming 128-tile deployment of the MWA—a data set that cannot be fully utilized using only the LT method. Our forecast predicts that 1000 hours of MWA observation should be enough to detect the fiducial 10 mK signal across much of the $k_\parallel$-$k_\perp$ plane accessible to the instrument. Moreover, we predict that the horizontal error bars on each band power estimate will be narrow, allowing each estimate to probe only a small range of scales.

Our results suggest several avenues for further research. Of course, the most immediate application is to begin analyzing the data already being produced by interferometers like LOFAR, GMRT, MWA, and PAPER as they start accumulating the sensitivity necessary to zero in on a detection of the EoR. The large volume of data these instruments promise to produce might make it useful to explore ways of further speeding up the Monte Carlo estimation of the Fisher matrix. There is significant redundancy in our calculated Fisher matrix because the window function shapes vary only relatively slowly with k-scale. We believe that one can reduce the number of Monte Carlo simulations needed to attain the same accuracy by adding a postprocessing step that fits the Fisher matrix to a parametrized form. This should work best in the regions of the $k_\parallel$-$k_\perp$ plane that are fairly uncontaminated by foregrounds, where Fisher matrix elements are expected to vary most smoothly. It may also be possible to speed up the Monte Carlo estimation of the Fisher matrix using the trace evaluation technology of [175].

The forecasting power of our method to see whether a particular observing campaign might reveal a particular aspect the power spectrum need not be limited to measurements of the EoR. Our method provides an opportunity to precisely predict what kind of measurement, and what kind of instrument, might be necessary for observing 21 cm brightness temperature fluctuations during the cosmic dark ages. Our method should prove useful for weighing a number of important design considerations: What is the optimal array configuration? What is the optimal survey volume? What about angular resolution? Spectral resolution? And in what sense are these

choices optimal for doing astrophysics and cosmology?

To help answer such questions, our technique could be used to compare the myriad of ideas for and possible implementations of future projects like HERA and the SKA and even to help find an optimal proposal. For example, one plan for achieving large collecting area is building a hierarchically regular array (a so-called "Omniscope") that takes advantage of FFT correlation [211] and redundant baseline calibration [124]. There exist many array configurations that fit into this category and it is not obvious what the optimal Omniscope might look like.

The quest to detect a statistical signal from the Epoch of Reionization is as daunting as it is exciting. It is no easy task to find that needle in a haystack of noise and foregrounds. However, now that we are for the first time armed with a method that can extract all the cosmological information from a massive data set without a prohibitive computational cost, we can feel confident that a sufficiently sensitive experiment can make that first detection—not just in theory, but also in practice.

2.A   Appendix: Toeplitz Matrices

In this appendix, we briefly review how to rapidly multiply by Toeplitz matrices. We need to employ the advantages of Toeplitz matrices because the assumption that our covariance matrices are diagonal in real space or in Fourier space, as was the case in [175], break down for covariance matrices with coherence lengths much larger than the box size.

A "Toeplitz" matrix is any matrix with the same number for every entry along its main diagonal and with every other diagonal similarly constant [80]. In general, a Toeplitz matrix is uniquely defined by the entries in its first row and its first column: if $i \geq j$ then $T_{ij} = T_{1+i-j,1}$ and if $i \leq j$ then $T_{ij} = T_{1,1-i+j}$. If the first row of a matrix is repeated with a cyclic permutation by one to the right in each successive row, then it is a special kind of Toeplitz matrix called a "circulant" matrix. Circulant matrices are diagonalized by the discrete Fourier transform [80]. Given a circulant matrix $\mathbf{C}$ with first column $\mathbf{c}$, the product of $\mathbf{C}$ and some arbitrary vector $\mathbf{v}$ can be computed

in $\mathcal{O}(N \log N)$ time because

$$\mathbf{Cv} = \mathbf{F}^{\dagger} \, \text{diag}(\mathbf{Fc}) \, \mathbf{Fv}, \tag{2.67}$$

where $\mathbf{F}$ is the unitary, discrete Fourier transform matrix [80]. Reading Equation 2.67 from right to left, we see that every matrix operation necessary for this multiplication can be performed in $\mathcal{O}(N \log N)$ time or better.

Conveniently, any symmetric Toeplitz matrix can be embedded in a circulant matrix twice its size. Given a symmetric Toeplitz matrix $\mathbf{T}$, we can define another symmetric Toeplitz matrix $\mathbf{S}$ with an arbitrary constant along its main diagonal. If we specify that the rest of the first row (besides the first entry) is the reverse of the rest of the first row of $\mathbf{T}$ (again ignoring the first entry), the fact that the matrix is Toeplitz and symmetric completely determines the other entries. For example,

$$\text{if } \mathbf{T} = \begin{pmatrix} 5 & 3 & 2 \\ 3 & 5 & 3 \\ 2 & 3 & 5 \end{pmatrix}, \text{ then } \mathbf{S} = \begin{pmatrix} 0 & 2 & 3 \\ 2 & 0 & 2 \\ 3 & 2 & 0 \end{pmatrix}. \tag{2.68}$$

It is straightforward to verify that the matrix $\mathbf{C}$, defined as

$$\mathbf{C} \equiv \begin{pmatrix} \mathbf{T} & \mathbf{S} \\ \mathbf{S} & \mathbf{T} \end{pmatrix}, \tag{2.69}$$

is a circulant matrix. We can now can multiply $\mathbf{C}$ by a zero-padded vector so as to yield the product of the Toeplitz matrix and the original vector, $\mathbf{Tx}$, that we are looking for:

$$\mathbf{C} \begin{pmatrix} \mathbf{v} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{Tv} \\ \mathbf{Sv} \end{pmatrix}. \tag{2.70}$$

Therefore, we can multiply any Toeplitz matrix by a vector in $\mathcal{O}(N \log N)$.

2.B   Appendix: Noise Covariance Matrix Derivation

In this appendix, we derive the the form of $\mathbf{N}$, the noise covariance matrix, in Equation 2.57 by combining the form of $P^N(\mathbf{k}, \lambda)$, the noise power spectrum, in Equation 2.32 with Equation 2.33, which relates $\mathbf{N}$ to $P^N(\mathbf{k}, \lambda)$. To accomplish this, we simplify $P^N(\mathbf{k}, \lambda)$ into a form that is more directly connected to our data cube. We then approximate the integrals in Equation 2.33 by assuming that the uv-coverage is piecewise constant in cells corresponding to our Fourier space grid.[13]

To simplify $P^N(\mathbf{k}, \lambda)$, we first note that because the term $B(\mathbf{k}, \lambda)$ in Equation 2.32 represents the synthesized beam and is normalized to peak at unity, we can reinterpret the factor of $(f^{\text{cover}})^{-1}B^{-2}(\mathbf{k}, \lambda)$ as an inverted and normalized uv-coverage. When $f^{\text{cover}} = 1$, the array has uniform coverage. We want to replace the factor $(f^{\text{cover}})^{-1}B^{-2}(\mathbf{k}, \lambda)$ with a quantity directly tied our choice of pixelization of the uv-plane and written in terms of the simplest observational specification: the total time that baselines spend observing a particular uv-cell, $t(\mathbf{k}, \lambda)$. We already know that the noise power is inversely proportional to that time because more time yields more independent samples.

To relate $t^{-1}(\mathbf{k}, \lambda)$ to $(f^{\text{cover}})^{-1}B^{-2}(\mathbf{k}, \lambda)$, we want to make sure that the formula yields the same answer for peak density in the case of a complete coverage. In other words, we want to find the constant t_{max} such that

$$\frac{t_{\text{max}}}{t(\mathbf{k}, \lambda)} = (f^{\text{cover}})^{-1}B^{-2}(\mathbf{k}, \lambda). \tag{2.71}$$

The time spent in the most observed cell is related to the size of the cell in the uv-plane, the density of baselines in that cell, and the total integration time of the observation, τ. The cell size is determined by the pixelization of our data cube. We have divided each slice of our data cube of size $L_x \times L_y$ into $n_x \times n_y$ pixels. In Fourier

---

[13]We also assume that measurements in nearby uv-cells are uncorrelated, which may not be true if the baselines are not coplanar; instead $\mathbf{N}$ would have to be modeled as sparse rather than diagonal in angular Fourier space.

space, this corresponds a pixel size of

$$\Delta k_x = \frac{2\pi}{L_x} = \frac{2\pi}{d_M \Delta\theta_x n_x}, \tag{2.72}$$

where $\Delta\theta_x$ and is the angular pixelization in the x direction. An equivalent relation is true for the y direction. Since $\Delta u = \Delta k_x d_M/(2\pi)$, we have that the area in uv-space of each of our grid points is

$$\Delta v \Delta u = \frac{1}{\Delta\theta_x n_x}\frac{1}{\Delta\theta_y n_y} = \frac{1}{\Omega_{\mathrm{pix}} n_x n_y}. \tag{2.73}$$

The maximum density of baselines is the density of the autocorrelations,[14] which is

$$n_{\max} = N_{\mathrm{ant}}\left(\frac{A_{\mathrm{ant}}}{\lambda^2}\right)^{-1}, \tag{2.74}$$

where the quantity $(A_{\mathrm{ant}}/\lambda^2)$ is the area in the uv-plane associated with a single baseline [139]. We thus have that

$$t_{\max} \equiv n_{\max}\Delta u \Delta v \tau = \frac{N_{\mathrm{ant}}\lambda^2 \tau}{A_{\mathrm{ant}}\Omega_{\mathrm{pix}} n_x n_y}. \tag{2.75}$$

Now we can substitute Equation 2.71 into Equation 2.32 to get a more useful form of $P^N(\mathbf{k}, \lambda)$:

$$P^N(\mathbf{k}, \lambda) = \frac{\lambda^4 T_{\mathrm{sys}}^2 y d_M^2}{A_{\mathrm{ant}}^2 \Omega_{\mathrm{pix}} n_x n_y}\frac{1}{t(\mathbf{k}_\perp, \lambda)}. \tag{2.76}$$

In general, $t(\mathbf{k}_\perp, \lambda)$ depends in a nontrivial way on the array layout. As such, the integral expression for $\mathbf{N}$ in Equation 2.33 with this form of $P^N(\mathbf{k}, \lambda)$ is only

---

[14]If the use of autocorrelations (which most observations throw out, due to their unfavorable noise properties) is troubling, then it is helpful to recall that for a large and fully-filled array, the uv-density of the shortest baselines is approximately the same as the uv-density of the autocorrelations.

analytically tractable along the line of sight. Integrating k_z, we get that

$$N_{ij} = \frac{\delta_{z_i z_j}}{\Delta z} \frac{\lambda^4 T_{\text{sys}}^2 y d_M^2}{A_{\text{ant}}^2 \Omega_{\text{pix}} n_x n_y} \int j_0^2(k_x \Delta x/2) j_0^2(k_y \Delta y/2) \times$$
$$e^{ik_x(x_i - x_j) + ik_y(y_i - y_j)} \frac{1}{t(\mathbf{k}_\perp, \lambda_i)} \frac{dk_x dk_y}{(2\pi)^2}. \tag{2.77}$$

We note that $\mathbf{N}$ is uncorrelated between frequency channels, as we would expect.

Along the other two dimensions, we will approach the problem by approximating the integrand as piecewise constant in Fourier cells, turning the integral into a sum and the dk into a Δk. We will use the index l to run over all Fourier modes perpendicular to the line of sight. Using the fact that the line of sight voxel length $\Delta z = y \Delta \nu$ and that $L_x L_y = \Omega_{\text{pix}} d_M^2 n_x n_y$, we have that

$$N_{ij} = \frac{\lambda^4 T_{\text{sys}}^2 \delta_{z_i z_j}}{A_{\text{ant}}^2 (\Omega_{\text{pix}} n_x n_y)^2 \Delta \nu} \sum_l^{\text{all x \& y}} \left[j_0^2(k_{x,l} \Delta x/2) \times \right.$$
$$\left. j_0^2(k_{y,l} \Delta y/2) e^{ik_{x,l}(x_i - x_j)} e^{ik_{y,l}(y_i - y_j)} \frac{1}{t_l(\lambda_i)} \right]. \tag{2.78}$$

Next, we can turn this form into one that is more clearly computationally easy to multiply by a vector by introducing another Kronecker delta:

$$N_{ij} = \frac{\lambda^4 T_{\text{sys}}^2 \delta_{z_i z_j}}{A_{\text{ant}}^2 (\Omega_{\text{pix}} n_x n_y)^2 \Delta \nu} \sum_l^{\text{all x \& y}} e^{ik_{x,l} x_i} e^{ik_{y,l} y_i}$$
$$\sum_m^{\text{all x \& y}} \left[j_0^2(k_{x,l} \Delta x/2) j_0^2(k_{y,l} \Delta y/2) e^{-ik_{x,m} x_j} e^{-ik_{y,m} y_j} \frac{\delta_{lm}}{t_l(\lambda_i)} \right]. \tag{2.79}$$

Finally, if we extend l and m to index over all frequency channels and all Fourier modes perpendicular to the line of sight, we can write down the noise covariance matrix as $\mathbf{N} = \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}} \mathbf{F}_\perp$ where $\mathbf{F}_\perp$ and $\mathbf{F}_\perp^\dagger$ are the discrete, unitary 2D Fourier and inverse Fourier transforms and where $\widetilde{\mathbf{N}}$ can be written as

$$\widetilde{N}_{lm} = \frac{\lambda^4 T_{\text{sys}}^2 j_0^2(k_{x,l} \Delta x/2) j_0^2(k_{y,l} \Delta y/2)}{A_{\text{ant}}^2 (\Omega_{\text{pix}})^2 n_x n_y \Delta \nu} \frac{\delta_{lm}}{t_l}. \tag{2.80}$$

The result, therefore, is a matrix that can be multiplied by a vector in $\mathcal{O}(N \log N)$.

2.C Appendix: Construction of the Preconditioner

In this final appendix, we show how to construct the preconditioner that we use to speed up the conjugate gradient method for multiplying $\mathbf{C}^{-1}$ by a vector. We devise our preconditioner by looking at $\mathbf{C}$ piece by piece, building up pairs of matrices that make our covariances look more like the identity matrix. We start with $\mathbf{C} = \mathbf{N}$, generalize to $\mathbf{C} = \mathbf{U} + \mathbf{N}$, and then finally incorporate $\mathbf{R}$ and $\mathbf{G}$ to build the full preconditioner.

2.C.1 Constructing a Preconditioner for N

Our first task is to find a pair of preconditioning matrices that turn $\mathbf{N}$ into the identity:

$$\mathbf{P_N}\mathbf{N}\mathbf{P_N^\dagger} = \mathbf{I}. \tag{2.81}$$

Because $\mathbf{N} = \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}} \mathbf{F}_\perp$, and because $\widetilde{\mathbf{N}}$ is a diagonal matrix, we define $\mathbf{P_N}$ and $\mathbf{P_N^\dagger}$ as follows:

$$\mathbf{P_N} = \widetilde{\mathbf{N}}^{-1/2}\mathbf{F}_\perp,$$
$$\mathbf{P_N^\dagger} = \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}}^{-1/2}. \tag{2.82}$$

Since applying $\mathbf{P_N}$ only requires multiplying by the inverse square root of a diagonal matrix and Fourier transforming in two dimensions, the complexity of applying $\mathbf{P_N}$ to a vector is less than $\mathcal{O}(N \log N)$.

2.C.2 Constructing a Preconditioner for U

The matrix $\mathbf{U}$ (Equation 2.55) can be written as the tensor product of three Toeplitz matrices, one for each dimension, bookended by two diagonal matrices, $\mathbf{D_U}$. Furthermore, since $\mathbf{D_U}$ depends only on frequency (as we saw in Section 2.3.4.2), its effect can be folded into $\mathbf{U}_z$ such that

$$\mathbf{D_U}[\mathbf{U}_x \otimes \mathbf{U}_y \otimes \mathbf{U}_z]\mathbf{D_U} \equiv \mathbf{U}_x \otimes \mathbf{U}_y \otimes \mathbf{U}_z'. \tag{2.83}$$

It is generally the case that $\mathbf{U}_x$ and $\mathbf{U}_y$ are both well approximated by the identity matrix. This reflects the fact that the spatial clustering of unresolved point sources is comparable with the angular resolution of the instrument. This assumption turns out to be quite good for fairly compact arrays, since for an array with 1 km as its longest baseline—the sort of compact array thought to be optimal for 21 cm cosmology—we expect an angular resolution on the order of 10 arcminutes, which is comparable to the fiducial value of 7 arcminutes that LT took to describe the clustering length scale for unresolved point sources. That value appears to be fairly reasonable given the results of [18, 77]. For the purposes of devising a preconditioner only, we can therefore adopt the simplification that

$$\mathbf{U} \approx \mathbf{I}_x \otimes \mathbf{I}_y \otimes \mathbf{U}_z, \tag{2.84}$$

where we have dropped the prime for notational simplicity. Looking back at Figure 2-4, this form of $\mathbf{U}$ neatly explains the stair-stepping behavior of the eigenvalues: for every eigenvalue of $\mathbf{U}_z$, $\mathbf{U}$ has $n_x \times n_y$ similar eigenvalues.

Since only a few eigenvalues of $\mathbf{U}_z$ are large, it is pedagogically useful to first address a simplified version of the preconditioning problem where $\mathbf{U}_z$ is approximated as a rank 1 matrix by cutting off its spectral decomposition after the first eigenvalue. We will later return to include the other relevant eigenvalues. We therefore write $\mathbf{U}$ as follows:

$$\mathbf{U} \approx \mathbf{I}_x \otimes \mathbf{I}_y \otimes \lambda \mathbf{v}_z \mathbf{v}_z^\dagger. \tag{2.85}$$

where $\mathbf{v}_z$ is the normalized eigenvector of $\mathbf{U}$.

Let us now take a look at the action of $\mathbf{P_N}$ and $\mathbf{P_N^\dagger}$ on $\mathbf{U} + \mathbf{N}$:

$$\begin{aligned}
\mathbf{P_N}(\mathbf{U} + \mathbf{N})&\mathbf{P_N^\dagger} \\
&= \mathbf{I} + \widetilde{\mathbf{N}}^{-1/2}\mathbf{F}_\perp(\mathbf{I}_x \otimes \mathbf{I}_y \otimes \lambda \mathbf{v}_z \mathbf{v}_z^\dagger)\mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}}^{-1/2} \\
&= \mathbf{I} + \widetilde{\mathbf{N}}^{-1/2}(\mathbf{I}_x \otimes \mathbf{I}_y \otimes \lambda \mathbf{v}_z \mathbf{v}_z^\dagger)\widetilde{\mathbf{N}}^{-1/2} \\
&\equiv \mathbf{I} + \overline{\mathbf{U}}. \tag{2.86}
\end{aligned}$$

Our next goal, therefore, is to come up with a new matrix $\mathbf{P_U}$ that, when applied to $\mathbf{I} + \overline{\mathbf{U}}$ gives us something close to $\mathbf{I}$.

We now take a closer look at $\overline{\mathbf{U}}$. Since it is a good approximation to say that $\widetilde{\mathbf{N}}$ only changes perpendicular to the line of sight,[15] we can rewrite $\overline{\mathbf{U}}$:

$$\begin{aligned}
\overline{\mathbf{U}} &\approx (\widetilde{\mathbf{N}}_\perp^{-1/2} \otimes \mathbf{I}_z)(\mathbf{I}_x \otimes \mathbf{I}_y \otimes \lambda \mathbf{v}_z \mathbf{v}_z^\dagger)(\widetilde{\mathbf{N}}_\perp^{-1/2} \otimes \mathbf{I}_z) \\
&= (\widetilde{\mathbf{N}}_\perp^{-1}) \otimes (\lambda \mathbf{v}_z \mathbf{v}_z^\dagger),
\end{aligned} \tag{2.87}$$

where $\widetilde{\mathbf{N}}_\perp$ is still a diagonal matrix, though only in two dimensions, generated from a baseline distribution averaged over frequency slices. We now form a pair of preconditioning matrices, $\mathbf{P_U}$ and $\mathbf{P_U^\dagger}$ of the form $\mathbf{P_U} \equiv \mathbf{I} - \beta\mathbf{\Pi}$ where $\mathbf{\Pi}$ has the property that $\mathbf{\Pi}\overline{\mathbf{U}} = \overline{\mathbf{U}}$ and that $\overline{\mathbf{U}}\mathbf{\Pi}^\dagger = \overline{\mathbf{U}}$. The matrix that fits this description is:

$$\begin{aligned}
\mathbf{\Pi} &= \widetilde{\mathbf{N}}^{-1/2}(\mathbf{I}_x \otimes \mathbf{I}_y \otimes \mathbf{v}_z \mathbf{v}_z^\dagger)\widetilde{\mathbf{N}}^{1/2} \\
&\approx (\mathbf{I}_x \otimes \mathbf{I}_y \otimes \mathbf{v}_z \mathbf{v}_z^\dagger) = \frac{1}{\lambda}\mathbf{U},
\end{aligned} \tag{2.88}$$

since $\widetilde{\mathbf{N}}$ only affects the x and y components and thus passes through the inner matrix. This also means that $\mathbf{\Pi} = \mathbf{\Pi}^\dagger$ and that $\mathbf{\Pi} = \mathbf{\Pi}^2$. The result for $\mathbf{P_U}(\mathbf{I} + \overline{\mathbf{U}})\mathbf{P_U^\dagger}$ is

$$(\mathbf{I} - \beta\mathbf{\Pi})(\mathbf{I} + \overline{\mathbf{U}})(\mathbf{I} - \beta\mathbf{\Pi}^\dagger) = \mathbf{I} + \overline{\mathbf{U}} - 2\beta\overline{\mathbf{U}} - 2\beta\mathbf{\Pi} + \beta^2\overline{\mathbf{U}} + \beta^2\mathbf{\Pi}. \tag{2.89}$$

The trick now is that for each uv-cell, $\overline{\mathbf{U}}$ has only one eigenvalue, which we call $\overline{\lambda}_l$ (again using l as an index over both directions perpendicular to the line of sight):

$$\overline{\lambda}_l = \frac{\lambda}{(\widetilde{N}_\perp)_{ll}}. \tag{2.90}$$

Likewise, $\mathbf{\Pi}$ only has one eigenvalue: 1. By design, these eigenvalues correspond to the same eigenvector. Since our goal is to have the matrix in equation (2.89) be the

---

[15]Were it not for the breathing of the synthesized beam with frequency, $\widetilde{\mathbf{N}}$, would only change perpendicularly to the line of sight. Since it is a small effect when considered over a modest redshift range, we can ignore it in the construction of our preconditioner. After all, we only need to make $\mathbf{PCP}^\dagger$ close to $\mathbf{I}$.

identity, we need only pick β such that:

$$1 = 1 + \overline{\lambda}_l - 2\beta\overline{\lambda}_l - 2\beta + \beta^2\overline{\lambda}_l + \beta^2. \tag{2.91}$$

Solving the quadratic equation, we get

$$\mathbf{P_U} \equiv \mathbf{I} - \left[\sum_l \left(1 - \sqrt{\frac{1}{1+\overline{\lambda}_l}} \right) \mathbf{v}_z \mathbf{v}_z^\dagger \otimes \overset{\leftrightarrow}{\delta}_{x,x_l} \otimes \overset{\leftrightarrow}{\delta}_{y,y_l} \right], \tag{2.92}$$

where the pair of $\overset{\leftrightarrow}{\delta}$ matrices pick out a particular uv-cell. If we want to generalize to more eigenvectors of $\mathbf{U}_z$, we simply need to keep subtracting off sums of matrices on the right hand side of Equation (2.92):

$$\mathbf{P_U} \equiv \mathbf{I} - \sum_k \left[\sum_l \left(1 - \sqrt{\frac{1}{1+\overline{\lambda}_{l,k}}} \right) \mathbf{v}_{z_k} \mathbf{v}_{z_k}^\dagger \otimes \overset{\leftrightarrow}{\delta}_{x,x_l} \otimes \overset{\leftrightarrow}{\delta}_{y,y_l} \right], \tag{2.93}$$

This works because every set of vectors corresponding to a value of k is orthogonal to every other set. Each term in the above sum acts on a different subspace of $\mathbf{C}$ independent of all the other terms in the sum.

If the relevant vectors $\mathbf{v}_{z_k}$ are precomputed, applying $\mathbf{P_U}$ can be done in $\mathcal{O}(Nm(\mathbf{U}_z))$ where $m(\mathbf{U}_z)$ is defined as the number of relevant eigenvalues of $\mathbf{U}_z$ that need preconditioning or, equivalently, the number of "steps" in the eigenvalues of $\mathbf{U}$ in Figure 2-4 above the noise floor. We examine how $m(\mathbf{U}_z)$ scales with the size of the data cube in Section 2.3.5.2. Because the fall off of the eigenvalues is exponential [121] we expect the scaling of m to be logarithmic.

In general, we can pick some threshold $\theta \geq 1$ to compare to the largest value of $\overline{\lambda}_{l,k}$ for a given k and then do not precondition modes with eigenvalues smaller than θ. One might expect there to be diminishing marginal utility to preconditioning the eigenvalues nearest 1. We explore how to optimally cut off the spectral decomposition in Section 2.3.5.3 by searching for a value of θ where the costs and benefits of preconditioning equalize.

2.C.3 Constructing a Preconditioner for R and G

We now turn our attention to the full matrix $\mathbf{C}$. The fundamental challenge to preconditioning all of the matrices in $\mathbf{C}$ simultaneously is that the components of $\mathbf{R}$ and $\mathbf{G}$ perpendicular to the line of sight are diagonalized in completely different bases. However, $\mathbf{U}$, $\mathbf{G}$, and $\mathbf{R}$ have very similar components parallel to the line of sight, due to the fact they all represent spectrally smooth radiation of astrophysical origin.

We can write down $\mathbf{R}$ as follows:

$$\mathbf{R} = \sum_n \left[\overset{\leftrightarrow}{\delta}_{x,x_n} \otimes \overset{\leftrightarrow}{\delta}_{y,y_n} \otimes \left(\sum_k \lambda_{n,k} \mathbf{v}_{z_{n,k}} \mathbf{v}_{z_{n,k}}^\dagger \right) \right], \tag{2.94}$$

which can be interpreted as a set of matrices describing spectral coherence, each localized to one point source, and all of which are spatially uncorrelated. And likewise, we can write down $\mathbf{G}$ as:

$$\mathbf{G} = \sum_{i,j,k} \left[\lambda_{x_i} \lambda_{y_j} \lambda_{z_k} \mathbf{v}_{x_i} \mathbf{v}_{x_i}^\dagger \otimes \mathbf{v}_{y_j} \mathbf{v}_{y_j}^\dagger \otimes \mathbf{v}_{z_k} \mathbf{v}_{z_k}^\dagger \right]. \tag{2.95}$$

We now make two key approximations for the purposes of preconditioning. First, we assume that all the z_k eigenvectors are the same, so $\mathbf{v}_{z_k} \approx \mathbf{v}_{z_{n,k}}$ for all n, all of which are also taken to be the same as the eigenvectors that appear in the preconditioner for $\mathbf{U}$ in Equation 2.93. Second, as in Section 2.C.2, we are only interested in acting upon the largest eigenvalues of $\mathbf{R}$ and $\mathbf{G}$. To this end, we will ultimately only consider the largest values of $\lambda_{n,k}$ and $\lambda_{i,j,k} \equiv \lambda_{x_i} \lambda_{y_j} \lambda_{z_k}$, which will vastly reduce the computational complexity of the preconditioner.

Our strategy for overcoming the difficulty of the different bases is to simply add the two perpendicular parts of the matrices and then decompose the sum into its eigenvalues and eigenvectors. We therefore define

$$\mathbf{\Gamma} \equiv \mathbf{R} + \mathbf{G} \tag{2.96}$$

(choosing the symbol $\boldsymbol{\Gamma}$ because it looks like $\mathbf{R}$ and sounds like $\mathbf{G}$). Given the above approximations, we can reexpress $\boldsymbol{\Gamma}$ as follows:

$$\boldsymbol{\Gamma} \approx \sum_k \left(\boldsymbol{\Gamma}_{\perp,k} \otimes \mathbf{v}_{z_k} \mathbf{v}_{z_k}^\dagger \right), \tag{2.97}$$

where we have defined each $\boldsymbol{\Gamma}_{\perp,k}$ as:

$$\boldsymbol{\Gamma}_{\perp,k} \equiv \left(\sum_n \lambda_{n,k} \overset{\leftrightarrow}{\delta}_{x,x_n} \otimes \overset{\leftrightarrow}{\delta}_{y,y_n} \right) \left(\sum_{i,j} \lambda_{i,j,k} \mathbf{v}_{x_i} \mathbf{v}_{x_i}^\dagger \otimes \mathbf{v}_{y_j} \mathbf{v}_{y_j}^\dagger \right). \tag{2.98}$$

Due to the high spectral coherence of the foregrounds, only a few values of k need to be included to precondition for $\boldsymbol{\Gamma}$. Considering the limit on angular box size imposed by the flat sky approximation and the limit on angular resolution imposed by the array size, this should require at most a few eigenvalue determinations of matrices no bigger than about 10^4 entries on a side. Moreover, those eigenvalue decompositions need only be computed once and then only partially stored for future use. In practice, this is not a rate-limiting step, as we see in Section 2.3.5.2.

We now write down the eigenvalue decomposition of $\boldsymbol{\Gamma}$:

$$\boldsymbol{\Gamma} = \sum_k \left(\sum_l \lambda_{l,k} \mathbf{v}_{\perp,l} \mathbf{v}_{\perp,l}^\dagger \right) \mathbf{v}_{z_k} \mathbf{v}_{z_k}^\dagger. \tag{2.99}$$

Before we attack the general case, we assume that only one value of $\lambda_{l,k}$ is worth preconditioning—we generalize to the full $\mathbf{P}_{\boldsymbol{\Gamma}}$ later. We now know that if we have a matrix that looks like $\mathbf{I} + \overline{\mathbf{U}}$ we can make it look like $\mathbf{I}$. So can we take $\mathbf{I} + \overline{\mathbf{U}} + \overline{\boldsymbol{\Gamma}}$, where $\overline{\boldsymbol{\Gamma}} \equiv \mathbf{P_N} \boldsymbol{\Gamma} \mathbf{P_N^\dagger}$, and turn it into $\mathbf{I} + \overline{\mathbf{U}}$? Looking at $\overline{\boldsymbol{\Gamma}}$,

$$\begin{aligned} \overline{\boldsymbol{\Gamma}} &= \widetilde{\mathbf{N}}^{-1/2} \mathbf{F}_\perp \lambda_{\boldsymbol{\Gamma}} \mathbf{v} \mathbf{v}^\dagger \mathbf{F}_\perp^\dagger \widetilde{\mathbf{N}}^{-1/2} \\ &= \lambda_{\boldsymbol{\Gamma}} (\widetilde{\mathbf{N}}_\perp^{-1/2} \widetilde{\mathbf{v}}_\perp \widetilde{\mathbf{v}}_\perp^\dagger \widetilde{\mathbf{N}}_\perp^{-1/2}) \otimes \mathbf{v}_z \mathbf{v}_z^\dagger, \end{aligned} \tag{2.100}$$

where $\lambda_{\boldsymbol{\Gamma}}$ is the sole eigenvalue we are considering and where $\widetilde{\mathbf{v}}_\perp \equiv \mathbf{F}_\perp \mathbf{v}_\perp$.

Again, we will look at a preconditioner of the $\mathbf{P_\Gamma} = \mathbf{I} - \beta\mathbf{\Pi}$ where:

$$\mathbf{\Pi} \equiv \left(\widetilde{\mathbf{N}}_\perp^{-1/2}\widetilde{\mathbf{v}}_\perp\widetilde{\mathbf{v}}_\perp^\dagger\widetilde{\mathbf{N}}_\perp^{1/2}\right) \otimes \mathbf{v}_z\mathbf{v}_z^\dagger. \tag{2.101}$$

This time, the $\widetilde{\mathbf{N}}_\perp^{\pm 1/2}$ matrices do not pass through the eigenvectors to cancel one another out. We now exploit the spectral similarity of foregrounds and the fact that $\widetilde{\mathbf{v}}_\perp^\dagger\widetilde{\mathbf{v}}_\perp = \mathbf{v}_z^\dagger\mathbf{v}_z = 1$ to obtain

$$\mathbf{P_\Gamma}\overline{\mathbf{U}}\mathbf{P_\Gamma}^\dagger = \overline{\mathbf{U}} + \frac{\lambda_{\mathbf{U}}}{\lambda_{\mathbf{\Gamma}}}(\beta^2 - 2\beta)\overline{\mathbf{\Gamma}}. \tag{2.102}$$

This is very useful because it means that if we pick β properly, we can get the second term to cancel the $\overline{\mathbf{\Gamma}}$ terms we expect when we calculate the full effect of $\mathbf{P_\Gamma}$ and $\mathbf{P_N}$ on $\mathbf{N} + \mathbf{U} + \mathbf{\Gamma}$. Noting that the sole eigenvalue of $\overline{\mathbf{\Gamma}}$ is $\overline{\lambda}_{\mathbf{\Gamma}} \equiv \lambda_{\mathbf{\Gamma}}\widetilde{\mathbf{v}}_\perp^\dagger\widetilde{\mathbf{N}}_\perp^{-1}\widetilde{\mathbf{v}}_\perp$, we also define $\overline{\lambda}_{\mathbf{U}} \equiv \lambda_{\mathbf{U}}\widetilde{\mathbf{v}}_\perp^\dagger\widetilde{\mathbf{N}}_\perp^{-1}\widetilde{\mathbf{v}}_\perp$. Multiplying our preconditioner by our matrices, we see that the the equality of the single eigenvalues yields another quadratic equation for β:

$$1 + \overline{\lambda}_{\mathbf{U}} = 1 - 2\beta + \beta^2 + (\beta^2 - 2\beta + 1)\overline{\lambda}_{\mathbf{\Gamma}} + \overline{\lambda}_{\mathbf{\Gamma}}\frac{\lambda_{\mathbf{U}}}{\lambda_{\mathbf{\Gamma}}}(\beta^2 - 2\beta). \tag{2.103}$$

Solving, we finally have our $\mathbf{P_\Gamma}$ that acts on $\mathbf{I} + \overline{\mathbf{U}} + \overline{\mathbf{\Gamma}}$ and yields $\mathbf{I} + \overline{\mathbf{U}}$:

$$\mathbf{P_\Gamma} = \mathbf{I} - \left(1 - \sqrt{\frac{1 + \overline{\lambda}_{\mathbf{U}}}{1 + \overline{\lambda}_{\mathbf{U}} + \overline{\lambda}_{\mathbf{\Gamma}}}}\right)\left[\left(\widetilde{\mathbf{N}}_\perp^{-1/2}\widetilde{\mathbf{v}}_\perp\widetilde{\mathbf{v}}_\perp^\dagger\widetilde{\mathbf{N}}_\perp^{1/2}\right) \otimes \mathbf{v}_z\mathbf{v}_z^\dagger\right]. \tag{2.104}$$

Finally, generalizing to multiple eigenvalues and taking advantage of the orthonormality of the eigenvectors, we have

$$\mathbf{P_\Gamma} = \mathbf{I} - \sum_{k,m}\left[\left(1 - \sqrt{\frac{1 + \overline{\lambda}_{\mathbf{U}_k}}{1 + \overline{\lambda}_{\mathbf{U}_k} + \overline{\lambda}_{\mathbf{\Gamma}_{k,m}}}}\right)\left(\left(\widetilde{\mathbf{N}}_\perp^{-1/2}\widetilde{\mathbf{v}}_{\perp_m}\widetilde{\mathbf{v}}_{\perp_m}^\dagger\widetilde{\mathbf{N}}_\perp^{1/2}\right) \otimes \mathbf{v}_{z_k}\mathbf{v}_{z_k}^\dagger\right)\right]. \tag{2.105}$$

The result of this somewhat complicated preconditioner is a reduction of the condition number of the matrix to be inverted by many orders of magnitude (see Figure 2-5).

Lastly, we include Fourier transforms at the front and the back of the precon-ditioner, so that the result, when multiplied by a real vector, returns a real vector. Therefore, the total preconditioner we use for $\mathbf{C}$ is:

$$\mathbf{F}_{\perp}^{\dagger}\mathbf{P_U}\mathbf{P_\Gamma}\mathbf{P_N}(\mathbf{R}+\mathbf{U}+\mathbf{N}+\mathbf{G})\mathbf{P_N^{\dagger}}\mathbf{P_\Gamma^{\dagger}}\mathbf{P_U^{\dagger}}\mathbf{F}_{\perp}. \tag{2.106}$$

Chapter 3

Mapmaking for Precision 21 cm Cosmology

The content of this chapter was submitted to Physical Review D *on October 8, 2014 and published [61] as* Mapmaking for precision 21 cm cosmology *on January 6, 2015.*

3.1 Introduction

The prospect of directly probing the intergalactic medium (IGM) during the cosmic dark ages, through the "Cosmic Dawn" and culminating with the Epoch of Reionization (EoR) has generated tremendous excitement in 21 cm cosmology over the past few years. Not only could it provide the first direct constraints on the astrophysics of the first stars and galaxies, but it could make an enormous new cosmological volume accessible to tomographic mapping—enabling exquisitely precise new tests of ΛCDM [135]. For recent reviews, see e.g. [74, 154, 188, 127].

More recently, that excitement has translated into marked progress toward a statistical detection of the 21 cm signal in the power spectrum. The first generation of experiments, including the Low Frequency Array (LOFAR [76]), the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER [171]), the Giant Metrewave Radio Telescope (GMRT [166]), and the Murchison Widefield Array (MWA [220, 29]) have already begun their observing campaigns. Both PAPER [105]

and the MWA [59] have released upper limits on the 21 cm power spectrum across multiple redshifts. PAPER has already begun to use their results to constrain some models of the thermal history of the IGM [173].

Still, the observational and analytical challenges that lie ahead for the field are considerable. The sensitivity requirements for a detection of the 21 cm power spectrum necessitate large collecting areas and thousands of hours of observation across multiple redshifts [151, 25, 117, 87, 170]. Of no less concern is the fact that the cosmological signal is expected to be dwarfed by foreground contaminants—synchrotron radiation from our Galaxy and other radio galaxies—by four or more orders of magnitude in brightness temperature at the frequencies of interest [53, 107, 18, 182, 244, 109].

The problem of power spectrum estimation in the presence of foregrounds has been the focus on considerable theoretical effort over the past few years [123, 122, 108, 225, 120, 58]. Liu and Tegmark [120] adapted inverse-covariance-weighted quadratic estimator techniques developed for Cosmic Microwave Background [208] and galaxy survey [213] power spectrum analysis to 21 cm cosmology. Dillon et al. [58] showed how those methods, which nominally take $\mathcal{O}(N^3)$ steps, where N is the number of voxels in a 3D map or "data cube", could be accelerated to as fast as $\mathcal{O}(N \log N)$.

However, both of those works took as their starting point data cubes containing signal, foregrounds, and noise. Neither considered the important impact that an interferometer has, not just on the noise in our maps, but on the maps themselves. An instrument-convolved map or "dirty map" has fundamentally different statistical properties than the underlying sky and the effects of the instrument cannot in general be fully undone. Dillon et al. [59] discussed this problem approximately by assuming that point spread functions (PSFs) or "synthesized beams" depended only on frequency. Generally speaking, that is not true; PSFs are direction-dependent and typically not invertible. In this work, we relax the assumption that went into Liu and Tegmark [120] and Dillon et al. [58] while retaining the goals they strove for: minimal information loss, rigorously understood statistics, and well-controlled approximations that make the analysis computationally feasible.

For any near-future 21 cm measurement, interferometric maps are essentially an

intermediate data compression step. The ultimate goal is to turn time-ordered data coming from the instrument—namely, visibilities—into statistical measurements that constrain our models of astrophysics and cosmology. So why even bother making a map if we are only going to take Fourier transforms of it and look at power spectra? The answer to that question depends on which strategy we pursue for separating the cosmological signal from foregrounds. There are two major approaches, which we will review presently.

Over the last few years, it has been realized that a region of cylindrical Fourier space[1] should be essentially free of foreground contamination [50, 172, 230, 156, 89, 225, 218, 125, 126]. We call this region the "EoR window" (see Figure 4-1). Observations of the EoR window thus far have found it noise dominated [182, 59]. For slowly varying spectral modes (i.e. low $k_\parallel$), the edge of the window is set by a combination of the intrinsic spectral structure of foreground residuals and the spectral structure introduced by the instrument. Fundamentally, an interferometer is a chromatic instrument and the fact that the shape of its point spread functions depends on frequency creates complex spectral structure in 3D maps of intrinsically smooth foregrounds [125, 126].

Fortunately, there is a theoretical limit to the region of Fourier space where instrumentally induced spectral structure can contaminate the power spectrum. It is set by the delay associated with a source at the horizon (which is the maximum possible delay) for any given baseline [172]. This region of cylindrical Fourier space is known colloquially as "the wedge." Furthermore, we expect that most of the foreground emission should appear in the main lobe of the primary beam, setting a soft limit on foreground emission at lower $k_\parallel$ (see Figure 4-1).

The simplest approach to power spectrum estimation in the presence of foregrounds, and likely the most robust, is to simply excise the entire section of Fourier space that could potentially be foreground-dominated. This conservative approach

---

[1]Points in cylindrical or "2D" Fourier space are denoted by $k_\parallel$, modes along the light of sight, and $k_\perp$, modes perpendicular to the line of sight. Cylindrical Fourier space takes advantage of isotropy perpendicular to the line of sight while keeping modes along the line of sight separate, since they are measured in a fundamentally different way.
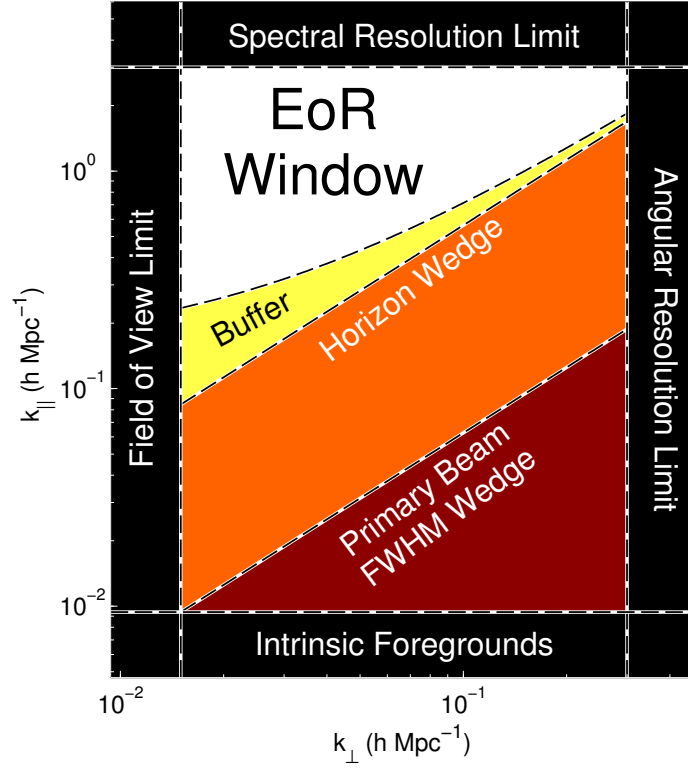
Figure 3-1: The "EoR window" is a region of Fourier space believed to be essentially foreground free and thus represents a major opportunity for detecting the 21 cm signal. Along the horizontal axis, the window is limited by the field of view, which sets the largest accessible modes, and the angular resolution of the instrument, which sets the smallest. Along the vertical axis, the window is limited by the spectral resolution of the instrument and by the intrinsic spectral structure of galactic and extragalactic foregrounds, which dominate the spectrally smooth modes. The EoR window is further limited by "the wedge," which results from the modulation of spectrally smooth foregrounds by the instrument's frequency-dependent and spatially varying point spread function. Much of the power in the wedge should fall below the wedge line associated with the primary beam while the horizon line serves as a hard cutoff for flat-spectrum foregrounds [172]. Limited "suprahorizon" emission has been observed and can be attributed to intrinsic spectral structure of the foregrounds [182], so it is possible we need a small buffer beyond the horizon to be certain that the window is foreground free. Without foreground subtraction, foregrounds are expected to dominate over the cosmological signal throughout the wedge.

144

takes the perspective that we have no knowledge about the detailed spatial or spectral structure of the foregrounds and therefore that the entire region under the wedge is hopelessly contaminated. If that were the case, the optimal strategy would simply be to project out those modes. This "foreground avoidance" strategy has been used to good effect by both PAPER [173, 105] and the MWA [59], though neither made sensitive enough measurements to be sure that foregrounds are sufficiently suppressed inside the EoR window to make a detection without subtracting them. Considerable work has already been done with methods of estimating the power spectrum that minimize foreground contamination from the wedge into the window [59, 126].

Foreground avoidance, however, comes at a significant cost to sensitivity. The more aggressive alternative is "foreground subtraction", a strategy that tries to remove power associated with foregrounds and expand the EoR window. The idea behind foreground subtraction is twofold. First, we remove our best guess as to which part of the data is due to foreground contamination. Second, we treat residual foregrounds as a form of correlated "noise," downweighting appropriately in the power spectrum estimator and taking into account biases introduced. In the limiting case where we know very little about the foregrounds, foreground subtraction becomes foreground avoidance.

For the upcoming Hydrogen Epoch of Reionization Array (HERA), Pober et al. [184] compared the effects of foreground avoidance to foreground subtraction. If the window can be expanded from delay modes associated with the horizon to delay modes associated with the full width at half maximum of the primary beam, the sensitivity to the EoR signal improves dramatically. Over one observing season with a 547-element HERA, the detection significance of a fiducial EoR signal improves from 38σ to 122σ. For smaller telescopes, this might mean the difference between an upper limit and a solid detection. More importantly, the errors on the measurements of parameters that describe reionization from the power spectrum improve from about 5% to less than 1% when employing extensive foreground subtraction. That would be the most sensitive measurement ever made of the direct effect of the first stars and galaxies on the IGM. Simply put, there is much that might be gained by an aggressive

foreground subtraction approach.

That said, it will not be easy. In order to expand the EoR window and reduce the effect of foregrounds, one must model them very carefully. Likely we will want to use outside information like high-resolution surveys to try to measure source fluxes to be much better than a percent. Even more importantly, one must take our own uncertainty about these models into account. If we do not, we risk mistakenly claiming a detection. We must propagate both our best estimates for the foregrounds and our uncertainty in our models through the instrument, which is the source of the wedge itself.

Both galactic and extragalactic foregrounds have complex spatial structure. Any precise model for their emission is direction dependent. More importantly, our model for the statistics of our uncertainty about their emission, is also direction dependent. The covariance of residual foregrounds, especially of bright sources, is most simply and compactly expressed in real space [58].

We can now finally answer the question of why we should make maps if we are ultimately interested in power spectra. We need maps as an intermediate data product because they allow us to prepare our data in a highly compressed form that puts us in a natural position to carefully pick apart the signal from the foregrounds and the noise. Forming power spectra directly with visibilities, by comparison, requires treating each local sidereal time separately and vastly increases the data volume. In Figure 3-2 we put mapmaking into the larger context of data reduction all the way from calibrated visibilities to cosmological and astrophysical constraints. The goal of each step is to reduce the volume of data while keeping as much cosmological information as possible, allowing for quantification of errors, and making the next step easier.

The science requirements of our maps are very different from those that motivated most interferometric mapmaking in radio astronomy to date. Usually, radio astronomers are interested in the astrophysics of what we call "foregrounds" and focus on detailed images and spectra. For us it is especially important to understand how our maps are related statistically to the true sky, whose underlying statistics we

146

Calibrated Visibilities

Mapmaking

Dirty Maps

Power Spectrum Estimation

Power Spectra

Parameter Estimation

Astrophysical & Cosmological Constraints
$\tau\ T_{\mathrm{vir}}$
$\Omega_\Lambda\ \Omega_k$
Σm_ν
$\Omega_b\ n_s\ \Omega_m$
$R_{\mathrm{mfp}}\ \alpha\ A_s\ \zeta$
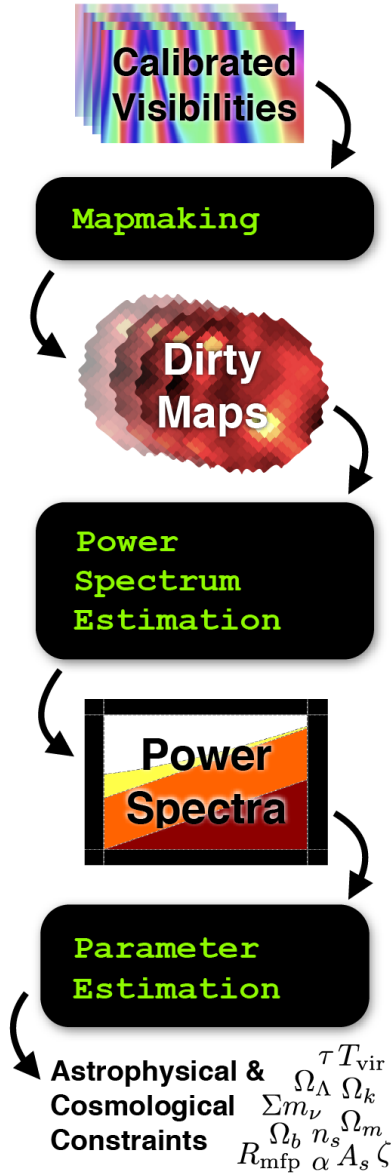
Figure 3-2: Mapmaking is the first in a series of steps that reduce the volume of data while trying not to lose any astrophysical or cosmological information. The goal of this work is to address that first data-compressional step—turning calibrated visibilities into a stack of dirty maps or a data cube—with any eye toward the next step—power spectum estimation in the presence of dominant astrophysical foregrounds. This data compression is achieved by combining together different observations a single, relatively small set of maps. Power spectra represent the cosmological signal even more compactly by taking advantage of homogeneity and isotropy and serve as the natural data product to connect to simulations and theory and thus constrain cosmological and astrophysical parameters.

would like to characterize using the power spectrum. Because interferometers do not uniformly or completely sample the Fourier plane, the relationship between our maps and the true sky is complicated. The PSFs of our maps depend both on frequency and on position on the sky. In order to estimate power spectra from maps accurately, we need to know precisely both the relationship of our dirty maps to the true sky and the covariance of our dirty maps that relates every pixel at every frequency to every other.[2] Current imaging techniques do not compute these quantities. It is the main point of this paper to show why and how that must be done.

Both [125] and [126] focused on a similar point about the important effect of the instrument on the power spectrum. There, the authors derived a framework for rigorously quantifying the errors and error correlations associated with instrument-convolved data and showed how the wedge feature arose even in a rigorous and optimal framework. However, because they formed power spectra directly from visibilities without using maps as an intermediate data-compression step, their tools are impractical for use with large data sets.

In this work, we have two main goals. First, we would like to mathematically understand how the instrument gives rise to a complicated PSF and how that PSF can be self-consistently incorporated into the inverse-covariance-weighted power spectrum estimation techniques (e.g. [120] and [58]). In Section 4.3.2, we discuss the theory of mapmaking as an intermediate step between observation and power spectrum estimation. Then, in Section 3.3, we investigate how to put that theory into practice. We use HERA as a case study in carrying out the calculation of dirty maps and their statistics. Although the computational cost of performing those calculations is naively quite large, we develop and analyze three main ways reducing it dramatically:

- We explore how restricting our maps to independent facets on the sky lets us reduce the number of elements in our PSF matrices and the difficulty of

---

[2]It is worth mentioning that the techniques developed here do not apply only to 21 cm tomography. Any power spectrum made with maps produced from interferometric data needs to take into account the effects of the frequency-dependent and spatially varying PSF on both the signal and the contaminants. This includes intensity mapping of CO and CII and interferometric measurements of the CMB. Higher-order statistics, like the bispectrum and trispectrum, also need precise knowledge of the relationship between the true sky and the dirty maps.

calculating them (Section 3.3.4).

- We show how individual timesteps can be combined and analyzed simultaneously, approximately accounting for the rotation of the sky over the instrument (Section 3.3.5).

- We show how the point spread functions, while not translationally invariant, vary smoothly enough spatially that the associated matrix operations can take advantage of certain symmetries for a computational speedup (Section 3.3.6).

We will show how each of these approximations works and analyze them to understand the trade-off between speed and accuracy in each case.

3.2 Precision Mapmaking And Map Statistics in Theory

Making maps from interferometric data has a long history and a great number of techniques have been developed with different science goals in mind [216]. Most focus on deconvolution, the removal of point source side lobes (or the side lobes of extended sources represented as multiple components) after their convolution with the synthesized beam. This is the basic idea behind the CLEAN algorithm [94] and its many descendants, including [196, 49, 48, 21, 35, 149, 19, 177, 201, 116, 203, 162]. Some of these, notably that of Sullivan et al. [203], take inspiration from [209], in that they use the framework of "optimal mapmaking" for forming dirty maps without losing any cosmological information contained in the visibilities. Additionally [197] and [198], which use the optimal mapmaking formalism in the m-mode basis to exploit the observational symmetries of a drift scanning interferometer, are also closely related to the work presented here.

A notable exception is [206], which develops a method of Bayesian deconvolution via Gibbs sampling in the relatively simplified case of a gridded uv-plane, which can then be used for power spectrum estimation [205]. This method not only calculates

a map but also gives error estimates on each pixel in that map. This is an especially promising technique for finding sources and quantifying the errors on our measurements of their fluxes and spectral indices. We take a different tack and do not focus on deconvolution at all.

In this work, we are interested not just in a dirty map but also in the statistical properties of that map. As in previous work, we want to know how sources are convolved with the instrument. But we also want to know how that instrumental convolution affects our covariance models for everything in the map, including signal, noise, and foregrounds. A complete understanding of the relationship between the true sky and our dirty maps will allow us to comprehensively model these important statistical quantities. Current imaging methods simply do not compute that relationship and the resulting noise covariance matrix. However, these are required for methods of power spectrum estimation in order to properly weight data in the presence of correlated noise and foregrounds and to account for missing modes. The importance of this was realized by [211], though we will use a different computational approach to speed up the calculations.

We begin this section by summarizing the relevant physics behind interferometry in Section 3.2.1. We then review the optimal mapmaking formalism in Section 3.2.2. Finally, in Section 3.2.3 we work out the consequences of proper map statistics for the inverse-covariance-weighted quadratic power spectrum estimation formalism, including how they affect the models of the covariance of cosmological signal, noise, and foreground residuals.

3.2.1   Interferometric Measurements

When we make maps from interferometric data, we are interested in computing a map estimator or "dirty map," which we call $\hat{\mathbf{x}}$, and understanding its relationship to $\mathbf{x}$, the true, discretized sky.[3] We do not have access to $\mathbf{x}$ directly; we can only make inferences about it by making a set of complex "visibility" measurements which we call

---

[3]We write these quantities as vectors as a compact way of combining indices over both angular dimensions on the sky and over frequency.

y. Each measurement made with our instrument is a linear combination of the true sky added to instrumental noise. Therefore, we can represent all our measurements with

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{A}$ represents the interferometric response of our instrument over all times, frequencies, and baselines and where each n_i is the instrumental noise on the ith visibility. The matrix $\mathbf{A}$ has the dimensions of the number of measured visibilities (for every baseline, frequency, and integration) by the number of voxels in the 3D sky (all pixels at all frequencies).

The statistics of $\mathbf{n}$ are fairly simple. It has zero mean and the noise on each visibility is generally treated as independent of that on every other visibility. Therefore,

$$\langle n_i \rangle = 0 \tag{3.2}$$

$$N_{ij} \equiv \langle n_i n_j^* \rangle = \sigma_i^2 \delta_{ij}. \tag{3.3}$$

The form of $\mathbf{A}$ is considerably more complicated, it can be written in the form of Equation (3.1) because a visibility is a weighted integral over the whole sky which can be approximated to any desired precision by a finite matrix operation.

The visibility measured by a noise-free instrument with arbitrarily fine frequency resolution at frequency ν and baseline $\mathbf{b}_m$ in response to a sky specific intensity $I(\hat{\mathbf{r}}, \nu)$ defined continuously over all points on the sky $\hat{\mathbf{r}}$ is

$$V(\mathbf{b}_m, \nu) = \int B_m(\hat{\mathbf{r}}, \nu) \, I(\hat{\mathbf{r}}, \nu) \exp\left[-2\pi i \frac{\nu}{c} \mathbf{b}_m \cdot \hat{\mathbf{r}}\right] d\Omega. \tag{3.4}$$

Here $B_m(\hat{\mathbf{r}}, \nu)$ is the product of the complex primary beams of the two antenna elements that form the mth baseline. In this equation and in the rest of this section, we will ignore the polarization of the sky and the fact that there are different beams for each polarization, assuming homogenous antenna elements. We do this for simplicity; the results are straightforwardly generalizable to a complete treatment of polarization, which we will explore in Appendix 3.A. In that appendix, we will also look at how

heterogenous arrays straightforwardly incorporated into our framework as well.

Given a finite number of measurements, we are interested in the relationship between visibilities and a discretized true sky, $\mathbf{x}$. In frequency, that discretization comes from the spectral response of our instrument—we can only measure a limited number of frequency channels. Spatially, we need to choose our pixelization of the sky. Let us define a 3D pixelization function $\psi_i(\hat{\mathbf{r}}, \nu)$ that incorporates both these kinds of pixelization. It is defined so that,

$$x_i = \int \psi_i(\hat{\mathbf{r}}, \nu) \frac{c^2}{2k_B\nu^2} I(\hat{\mathbf{r}}, \nu) d\Omega d\nu, \tag{3.5}$$

where the extra factor of $c^2/2k_B\nu^2$ converts from units of specific intensity to brightness temperature. For simplicity, we define $\psi_i(\hat{\mathbf{r}}, \nu)$ to be the unitless top-hat function, normalized such that

$$\int \psi_i(\hat{\mathbf{r}}, \nu) \frac{d\Omega}{\Delta\Omega} \frac{d\nu}{\Delta\nu} = 1 \tag{3.6}$$

where $\Delta\nu$ is the frequency resolution of the instrument and $\Delta\Omega$ is the angular size of the pixels. Other choices of $\psi_i(\hat{\mathbf{r}}, \nu)$ are perfectly acceptable, in which case $\Delta\nu$ and $\Delta\Omega$ become characteristic spectral and spatial sizes of pixels.

Therefore we can rewrite Equation (3.4) as a sum:

$$V(\mathbf{b}_m, \nu_n) \approx \sum_k \Delta\Omega \frac{2k_B\nu_n^2}{c^2} x_k(\nu_n) B_m(\hat{\mathbf{r}}_k, \nu_n) \exp\left[-2\pi i \frac{\nu_n}{c} \mathbf{b}_m \cdot \hat{\mathbf{r}}_k\right]. \tag{3.7}$$

Here we have chosen to break apart the index i into a spatial subindex, k, and a spectral subindex, n. The sum is over all spatial pixels. This approximation relies on choosing a frequency and angular resolution small enough that $B(\hat{\mathbf{r}}, \nu)$ and $\exp\left[-2\pi i (\nu/c)\mathbf{b}_m \cdot \hat{\mathbf{r}}\right]$ can be approximated as constants inside of a single spatial pixel and frequency channel. Since $V(\mathbf{b}_m, \nu_n)$ is an entry in $\mathbf{y}$, Equation (3.7) gives us the elements of $\mathbf{A}$ by relating $\mathbf{y}$ to $\mathbf{x}$ for a single observation and a single baseline. Of course, the full matrix $\mathbf{A}$ that goes into Equation (3.1) gives us a relationship between the true sky and every visibility at every frequency and at every local sidereal time. The basic physics, however, is captured by Equation (3.7).

3.2.2 The Optimal Mapmaking Formalism

Given a set of visibilities (or any time-ordered data) of the form in Equation (3.1), there is a well known technique for forming estimators of the true sky without losing any information about the discretized sky contained in the time-ordered data [209]. Those estimators, known as "optimal mapmaking" estimators, take the general form

$$\widehat{\mathbf{x}} = \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{y} \tag{3.8}$$

where $\mathbf{D}$ can be any invertible normalization matrix. Especially for long observations, $\mathbf{y}$ is a much larger vector than $\widehat{\mathbf{x}}$. Mapmaking represents a major data compression step.

The expected value of the estimator is

$$\begin{aligned} \langle\widehat{\mathbf{x}}\rangle &= \langle\mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}(\mathbf{A}\mathbf{x} + \mathbf{n})\rangle \\ &= \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}(\mathbf{A}\mathbf{x} + \langle\mathbf{n}\rangle) \\ &= \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A}\mathbf{x}. \end{aligned} \tag{3.9}$$

In general, the expected value of $\widehat{x}$ is not the same as the true sky but is rather some complicated linear combination of pixels on the true sky. We define

$$\mathbf{P} \equiv \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A} \tag{3.10}$$

to be the matrix of point spread functions. Each column of this matrix tells us how each pixel on the true sky gets mapped to all the pixels of the dirty map. If we want to normalize the PSF to always have a central value of 1, we can achieve that by a judicious choice of $\mathbf{D}$. In this work, we make that choice of PSF normalization. Recall that $\mathbf{D}$ can be any invertible matrix. Since we are not trying to make images that look as much as possible like the true sky but rather just to keep track of exactly how our dirty maps are related to the true sky, making a very simple choice for $\mathbf{D}$
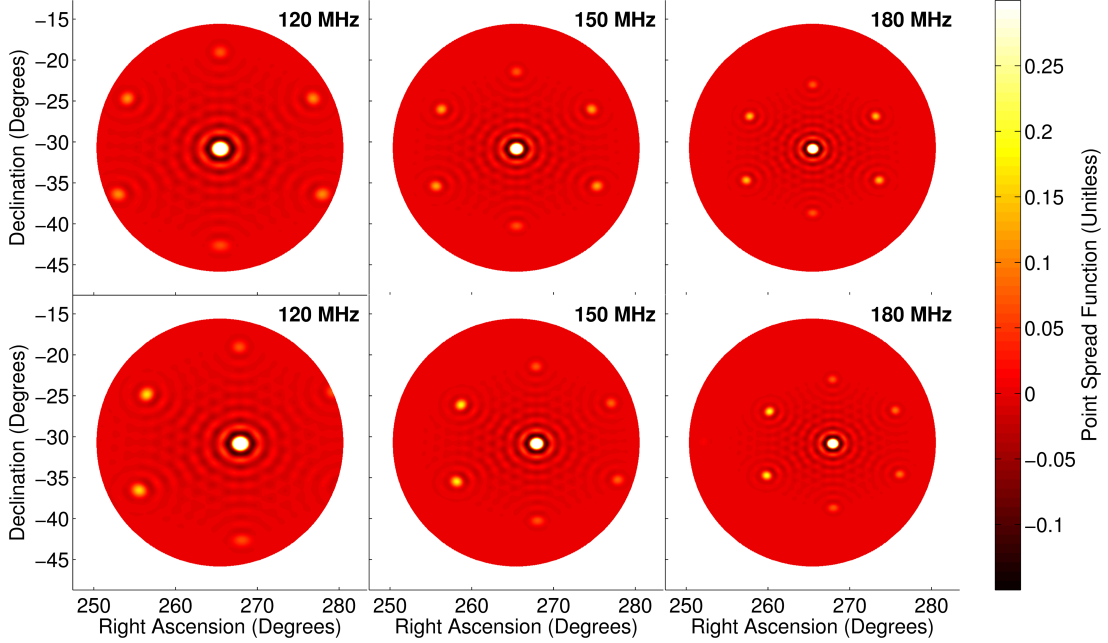
Figure 3-3: The point spread function (or equivalently, the synthesized beam) of a dirty map varies both as a function of position on the sky and as a function of frequency. In the top row, we show the point spread functions at three frequencies corresponding to the center of the primary beam calculated for HERA. They exhibit clear diffraction rings and fairly strong side lobes due to tje fact that the minimum separation between antennas is significantly longer than the wavelength. The hexagonal pattern is due to the geometry of the array. In the bottom row, we look at off-center point spread functions. These also have side lobes, though they are asymmetric due to the primary beam and the projected layout of the array and thus a clear example of the translational variation of the PSF. All six can be thought of as single rows of different frequency blocks of the full matrix of point spread functions, $\mathbf{P}$. Each PSF peaks at 1, but we have saturated the color scale to show detail. In Section 3.3, we will explain in detail how these PSFs are calculated.

is sensible.[4] Therefore, we use our freedom in choosing $\mathbf{D}$ to make it a diagonal matrix—effectively a per-pixel normalization. In Figure 3-3 we plot an example of the central portions of two different rows of $\mathbf{P}$ at three different frequencies.

---

[4]The choice of $\mathbf{D} = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1}$ was used by WMAP [106] because it makes $\mathbf{P} = \mathbf{I}$, but that matrix is generally not invertible in radio interferometry. Whenever one cannot make that choice of $\mathbf{D}$, $\mathbf{P}$ is not the identity and one must keep track of its effects.

3.2.3 Connecting Maps to Power Spectra

As we discussed earlier, we are interested in mapmaking in order to reduce the volume of our data without losing any sky information or the ability to remove foregrounds. From the map, the next step is to further compress the data by calculating a power spectrum, which can be directly compared with theoretical predictions. To connect the mapmaking formalism to 21 cm power spectrum estimation, we will review the statistical estimator formalism for calculating power spectra while not losing any cosmological information. In the process, we will enumerate the quantities that we need to calculate in order to estimate a power spectrum from $\widehat{\mathbf{x}}$. Then we will show the form that those quantities take in terms of $\widehat{\mathbf{x}}$, $\mathbf{P}$, and $\mathbf{D}$.

3.2.3.1 Power Spectrum Estimation Reivew

Fundamentally, a power spectrum estimate is a quadratic combination of the data. To calculate a power spectrum, roughly speaking, one simply Fourier transforms real-space data, squares, and then averages in discrete bins to form "band powers." In a real-world measurement with noise and foreground contamination, we need a more sophisticated technique.

Because we have a finite amount of data, we must discretize the power spectrum we estimate by approximating $P(\mathbf{k})$ as a piecewise constant function described by a set of band powers $\mathbf{p}$ using

$$P(\mathbf{k}) \approx \sum_{\alpha} p_{\alpha} \chi_{\alpha}(\mathbf{k}), . \tag{3.11}$$

Here $\chi_{\alpha}(\mathbf{k})$ is a characteristic function which equals 1 inside the region described by the band power p_{α} and vanishes elsewhere.

Since the power spectrum is a quadratic quantity in the data, an estimator $\widehat{\mathbf{p}}$ of the band power spectrum $\mathbf{p}$ (which is discretized by approximating the power spectrum as piecewise-constant) takes the form

$$\widehat{p}_{\alpha} = (\widehat{\mathbf{x}} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{E}_{\alpha} (\widehat{\mathbf{x}} - \boldsymbol{\mu}) - b_{\alpha}. \tag{3.12}$$

Here $\mathbf{E}_\alpha$ very generally represents the operations we want to perform on the data and $\boldsymbol{\mu} \equiv \langle \widehat{\mathbf{x}} \rangle$ is the ensemble average over many realizations of the same exact observation, each with different noise, and $\mathbf{b}$ removes additive bias from noise and residual foregrounds in the power spectrum.

Just as estimators of the form in Equation (3.8) do not lose any information about the true sky contained in the visibilities, there exists an optimal quadratic estimator for power spectra that does not lose cosmological information [208].[5] Those estimators take the form

$$\widehat{p}_\alpha = \frac{1}{2} M^{\alpha\beta} (\widehat{\mathbf{x}} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{C}^{-1} \mathbf{C}_{,\beta} \, \mathbf{C}^{-1} (\widehat{\mathbf{x}} - \boldsymbol{\mu}) - b_\alpha. \tag{3.13}$$

In this equation, $\mathbf{M}$ is an invertible normalization matrix, analogous to $\mathbf{D}$ and $\mathbf{C}$ is the covariance of $\widehat{\mathbf{x}}$ (not of the true sky $\mathbf{x}$) and is defined as

$$\mathbf{C} \equiv \langle \widehat{\mathbf{x}} \widehat{\mathbf{x}}^{\mathsf{T}} \rangle - \langle \widehat{\mathbf{x}} \rangle \langle \widehat{\mathbf{x}} \rangle^{\mathsf{T}}. \tag{3.14}$$

Each $\mathbf{C}_{,\beta}$ matrix, which encodes the Fourier transforming and binning steps of the power spectrum, is defined such that

$$\mathbf{C} = \mathbf{C}^{\text{contaminants}} + \sum_\beta p_\beta \mathbf{C}_{,\beta} \,. \tag{3.15}$$

Here $\mathbf{C}^{\text{contaminants}}$ represents the covariance of anything that appears in $\widehat{\mathbf{x}}$ that is not the 21 cm cosmological signal. In other words, the set of $\mathbf{C}_{,\beta}$ matrices tells us how the covariance of $\widehat{\mathbf{x}}$ responds to changes in the underlying band powers, $\mathbf{p}$. We will explain the precise form of $\mathbf{C}_{,\beta}$ shortly.

3.2.3.2 The Statistics of the Mapmaking Estimator

All of the quantities we are interested in calculating when estimating the power spectrum, including the bias term, the errors on our band powers, the error covariance

---

[5]This entails certain assumptions, most notably that the noise, residual foregrounds, and signal are all completely described by their means and covariances—in other words that they are Gaussian. We know that this is not exactly true in the case of residual foregrounds and signal, though it is generally assumed to be a pretty good approximation for the purposes of the first generation of 21 cm measurements [120].

between band powers, and the "window functions" that encode the relationship between $\widehat{\mathbf{p}}$ and $\mathbf{p}$, are derived from our models of $\boldsymbol{\mu}$ and $\mathbf{C}$ (see e.g. [208, 120, 58, 59] for the exact forms of these quantities). In this section, we will see how those quantities depend on the mapmaking algorithm and are inextricably linked to the response of the interferometer.

We have already shown that $\langle\widehat{\mathbf{x}}\rangle = \mathbf{Px}$ in Equations (3.9) and (3.10). When we are making a map, this is sufficient—there is a "true" sky and we are trying to estimate a quantity related to it from noisy data in a well-understood way. In the context of power spectrum estimation, simply averaging down instrumental noise is not enough. Because we are interested in the statistical properties of the Universe as a whole, we are trying to use multiple independent spatial modes to learn about at the underlying statistics of $\mathbf{x}$, taking advantage of homogeneity and isotropy. Though there is only one true sky, we treat it as a random field with Gaussian statistics. Therefore,

$$\boldsymbol{\mu} = \langle\widehat{\mathbf{x}}\rangle = \mathbf{P}\langle\mathbf{x}\rangle = \mathbf{P}\left[\langle\mathbf{x}^S\rangle + \langle\mathbf{x}^N\rangle + \langle\mathbf{x}^{FG}\rangle\right] = \mathbf{P}\langle\mathbf{x}^{FG}\rangle. \tag{3.16}$$

Here we have explicitly separated our model for the sky into three statistically independent parts: the 21 cm signal, the noise, and the foregrounds. Only the foregrounds have nonzero mean.[6] Because they are statistically independent, the covariance can be separated into the sum of three matrices.[7] Hence,

$$\mathbf{C} = \mathbf{C}^S + \mathbf{C}^N + \mathbf{C}^{FG}. \tag{3.17}$$

We will now show how all of these are calculated in the context of optimal mapmaking.

---

[6]The mean of the cosmological signal is zero only because it is usually defined as the fluctuations from the mean brightness temperature of the global 21 cm signal. For our purposes, the global signal is a contaminant and can be treated as part of the diffuse foregrounds without loss of generality.

[7]It should be noted that each of these covariance matrices is the covariance of the instrument-convolved sky and not the true sky, in contrast to the notation in [58] which, by treating an idealized scenario, ignored the distinction.

3.2.3.3   The Signal Covariance

First, let us turn to the signal covariance, $\mathbf{C}^S$. To understand what this really means, we need to first explain what we mean by $\mathbf{x}^S$. Imagine a continuous 21 cm temperature field as a function of position in comoving coordinates, $x^S(\mathbf{r})$. Each element of the vector $\mathbf{x}^S$ is given by

$$x_i^S \equiv \int \psi_i(\mathbf{r}) x^S(\mathbf{r}) \frac{d^3r}{\Delta V}, \tag{3.18}$$

where $\psi_i(\mathbf{r})$ encloses exactly the same volume as $\psi_i(\hat{\mathbf{r}}, \nu)$ and $\Delta V \equiv \int \psi_i(\mathbf{r}) d^3r$ is the comoving volume of a voxel. The continuous 21 cm power spectrum, $P(\mathbf{k})$ is defined by

$$\left\langle \left[\widetilde{x}^S(\mathbf{k})\right]^* \widetilde{x}^S(\mathbf{k}')\right\rangle \equiv (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}') P(\mathbf{k}), \tag{3.19}$$

where $\widetilde{x}^S(\mathbf{k})$ is the Fourier transform of $x^S(\mathbf{r})$. It follows then that

$$\langle x_i^S x_j^S \rangle - \langle x_i^S \rangle \langle x_j^S \rangle = \int \widetilde{\psi}_i(\mathbf{k}) \widetilde{\psi}_j^*(\mathbf{k}) P(\mathbf{k}) \frac{d^3k}{(2\pi)^3}. \tag{3.20}$$

By combining Equations (3.20) and (3.11), we can write down the covariance of $\mathbf{x}^S$:

$$\langle x_i^S x_j^S \rangle - \langle x_i^S \rangle \langle x_j^S \rangle \approx \sum_\alpha p_\alpha Q_{ij}^\alpha, \tag{3.21}$$

where

$$Q_{ij}^\alpha \equiv \int \widetilde{\psi}_i(\mathbf{k}) \widetilde{\psi}_j^*(\mathbf{k}) \chi_\alpha(\mathbf{k}) \frac{d^3k}{(2\pi)^3}. \tag{3.22}$$

Finally, using the fact that $\langle \widehat{\mathbf{x}} \rangle = \mathbf{P}\mathbf{x}$ determines also the relationship between the cosmological components of $\mathbf{x}$ and $\widehat{\mathbf{x}}$, we find that

$$\mathbf{C}^S \approx \mathbf{P} \left[\sum_\alpha p_\alpha \mathbf{Q}_\alpha \right] \mathbf{P}^\mathsf{T} \tag{3.23}$$

and therefore that

$$\mathbf{C}_{,\alpha} \approx \mathbf{P}\mathbf{Q}_\alpha \mathbf{P}^\mathsf{T}. \tag{3.24}$$

3.2.3.4 The Noise Covariance

While $\langle \widehat{\mathbf{x}}^N \rangle = \langle \mathbf{x}^N \rangle = 0$, the instrumental noise still contributes to the covariance. Our mapmaking formalism makes it straightforward to track how the noise on individual visibilities, σ_i^2, translates into correlated noise between pixels in a dirty map, which is described by $\mathbf{C}^N$. Let us imagine that $\mathbf{x} = 0$ and our instrument measured just noise for each visibility. If we compute the covariance of $\widehat{\mathbf{x}}$ in this case we will have $\mathbf{C}^N$, since $\mathbf{C}^S$ and $\mathbf{C}^{FG}$ represent our knowledge about the sky. This is true because there are no cross terms that correlate noise with foregrounds or signal.

Therefore, since our usual inverse-covariance-weighted map estimator now gives us

$$\widehat{\mathbf{x}}^N = \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{n}, \tag{3.25}$$

it follows that

$$\begin{aligned}
\mathbf{C}^N = \left\langle \widehat{\mathbf{x}}^N \left(\widehat{\mathbf{x}}^N\right)^\mathsf{T} \right\rangle &= \left\langle \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{n}\mathbf{n}^\dagger\mathbf{N}^{-1}\mathbf{A}\mathbf{D}^\mathsf{T} \right\rangle \\
&= \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1} \left\langle \mathbf{n}\mathbf{n}^\dagger \right\rangle \mathbf{N}^{-1}\mathbf{A}\mathbf{D}^\mathsf{T} \\
&= \mathbf{D}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A}\mathbf{D}^\mathsf{T} = \mathbf{P}\mathbf{D}^\mathsf{T}. \tag{3.26}
\end{aligned}$$

This is a gratifyingly simple result; calculating $\mathbf{P}$ yields $\mathbf{C}^N$ virtually for free. It also allows us to avoid the common assumption (made for example by [59], [120] and, [58]) that instrumental noise is uncorrelated between pixels in a gridded uv-plane. Correlations between uv pixels introduced by the primary beam are fully taken into account in our framework because, like in [125], $\mathbf{C}^N$ contains all the relevant information about the instrument and the mapmaking process.

3.2.3.5 The Foreground Covariance

Finally, we come to the statistics of the foregrounds. The reason that we treat $\mathbf{x}^{FG}$ as a random field even though there is really only one set of true foregrounds is that we want to represent both our best guess at the foregrounds and our uncertainty about that guess. When we write $\langle \mathbf{x}^{FG} \rangle$ in Equation (3.16), we really mean our best

guess as to the true foregrounds—the average of our incomplete knowledge about their positions, fluxes, spectral indices, and angular extents. Therefore we need to calculate

$$\boldsymbol{\mu} = \langle \widehat{\mathbf{x}}^{FG} \rangle = \mathbf{P} \langle \mathbf{x}^{FG} \rangle = \mathbf{P} \mathbf{x}^{FG}_{\text{model}} \tag{3.27}$$

to use in our quadratic estimator in Equation (3.13).

Previous work (e.g. [120, 58]) built explicit models of the foreground uncertainty by looking at the first and second moments of $\mathbf{x}^{FG}$ and not at $\widehat{\mathbf{x}}^{FG}$. We can take that work and generalize it straightforwardly. If $\mathbf{C}^{FG}_{\text{model}}$ is a model of foregrounds that takes into account our uncertainties about fluxes, spectral indices, and angular correlations, like the one developed in [120] and [58], then the foreground covariance of the estimator is

$$\mathbf{C}^{FG} = \mathbf{P} \mathbf{C}^{FG}_{\text{model}} \mathbf{P}^{\mathsf{T}}. \tag{3.28}$$

This equation compactly illustrates a key difference between the analysis methods developed by Liu and Tegmark [120] and Dillon et al. [58] and any future work that takes into account the inherent frequency dependence of foregrounds in dirty maps—the focus of this work. Intrinsic foregrounds are believed to be dominated by only a few Fourier modes [121]. That means that the expression of our uncertainty about the level of foreground contamination and thus our ability to subtract foreground, C^{FG}_{model}, should also be dominated by a few Fourier modes. However the PSF's spectral and spatial structure moves power from those low $k_{\parallel}$ modes up into the wedge. In Figure 3-4, we plot a few representative lines of sight of a field-centered PSF of a zenith-pointed instrument at different distances from field center. Even a flat-spectrum source would see considerable structure introduced on many spatial scales along the line of sight, especially far from the zenith. This is the origin of the wedge [156] and, as [125] pointed out, it can be fully understood as a consequence of the fact that frequency appears in the exponent of Equation (3.4). An interferometer is an inherently chromatic instrument.

To summarize, in order to optimally estimate a 21 cm power spectrum from the results of an optimal mapmaking routine, we must properly take into account the
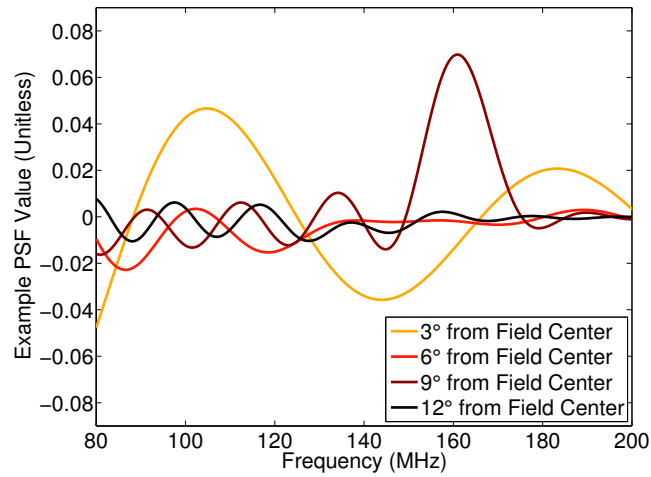
Figure 3-4: The position and frequency dependence of the synthesized beam is the origin of the "wedge" feature and plays a key role in determining which Fourier modes are foreground dominated in any power spectrum estimate. Here we show four different example lines of sight through a single frequency-dependent PSF, namely the one we showed for HERA in the top row of Figure 3-3. The structure we see means that intrinsically flat spectrum sources will appear far more complicated in a dirty map. We can also see that emission further from the zenith has more complicated spectral structure—an observation that helps explain the wedge. Any attempt at foreground subtraction will require detailed knowledge of this spectral behavior, both for our models for foregrounds and for our models of our uncertainty about foreground fluxes and spectral indices.

relationship between the dirty map and the true sky. To do this, we will need:

1. Our estimated dirty map, $\widehat{\mathbf{x}}$.

2. The normalization matrix for that map, $\mathbf{D}$, and the matrix of point spread functions, $\mathbf{P}$. Those require knowledge of the instrument, the observing strategy, and the noise in our measurements.

3. A model for the cosmological signal, which will allow us to properly account for sample variance.

4. A "best guess" for the foregrounds and a model for our uncertainty about that best guess.

With all these components, we can go from visibilities, through the data-compressing mapping step, and all the way to band powers in a self-consistent way while minimizing the loss of cosmological information and maintaining a full understanding of the error properties of our measurements.

3.3 Precision Mapmaking in Practice: Methods, Trade-Offs, and Results

The theoretically optimal mapmaking method outlined in Section 4.3.2 poses immense computational challenges. To make it useful for real-world application, we need to find and assess ways of simplifying it while maintaining its precision and statistical rigor.

Because this work serves in large part to generalize the work of [58], it is essential to continue to assess that the proposed algorithms are computationally feasible, despite the large size of these data sets and the potentially cost-prohibitive matrix operations involved. That work showed that as long as $\mathbf{C}$ could be decently preconditioned and then multiplied by a vector quickly, we could estimate the power spectrum in a way that scaled favorably with the data volume—between $\mathcal{O}(N \log N)$ and $\mathcal{O}(N^{5/3})$, where N is the number of voxels in a data volume. This was accomplished using

various numerical tricks, taking advantage of translational invariance, the fast Fourier transform, various symmetries, and the flat-sky approximation.

Without any approximations, the vectors and matrices we introduced in Section 4.3.2 are very big. $\mathbf{P}$, for example relates the whole true sky to the whole dirty map—for every frequency, it has as many entries as the number of pixels squared. The time-ordered data vector is very big too—it has entries for every baseline, at every frequency, for every integration. That means that $\mathbf{A}$ is enormous, since it maps from $\mathbf{x}$ to $\mathbf{y}$. We quantify exactly the exact scale of the problem of data volume and computational difficulty in Section 3.3.3, but it is clear that calculating every vector and matrix quantity we have enumerated in Section 4.3.2 is not feasible.

When making maps, there are at least six ways to make $\widehat{\mathbf{x}}$ and $\mathbf{P}$ smaller or easier to calculate or use. Three have to do with the geometry of $\widehat{\mathbf{x}}$; three have to do with approximate methods of calculating $\widehat{\mathbf{x}}$ or $\mathbf{P}$:

1. We can make faceted maps of only very small parts of the sky at a time.

2. We can pixelize the sky more coarsely.

3. We can average together neighboring frequencies, lowering the frequency resolution.

4. We can average together neighboring timesteps before computing $\mathbf{P}$.

5. We can make $\mathbf{P}$ smaller by taking advantage of the finite sizes of the primary and the synthesized beams.

6. We can make $\mathbf{P}$ sparser by approximately fitting it in some basis.

Roughly speaking, the first three approaches affect the kind of maps we want to make and the information content in them. The last three affect the quality of the maps we make or the fidelity with which an approximate version of $\mathbf{P}$ represents the relationship between $\widehat{\mathbf{x}}$ and $\mathbf{x}$. The exact properties of the desired maps depends upon the power spectrum estimation technique used. For example, if we want to measure high $k_\perp$ modes, we need high angular resolution and therefore a lot of pixels.

In this work, we take a specific case of the first three—choices motivated by the particular array we assess and the desire not to lose much cosmological information. We then evaluate quantitatively the trade-offs inherent in approaches that affect the quality of $\widehat{\mathbf{x}}$ and any approximation to $\mathbf{P}$. We begin by specifying both the array (Section 3.3.1) and the sky model (Section 3.3.2) that we use for the case study we present. In that context, we can quantify the computational challenges involved in mapmaking in Section 3.3.3.

From there, we examine the three ways of making the mapmaking problem easier for a given kind of map. In Section 3.3.4 we look at truncating $\mathbf{P}$ and how that affects our understanding of the relationship between the dirty map and the true sky. In Section 3.3.5 we look at the optimal way to perform time averaging and the trade-offs involved. Then we look at finding a sparse approximation to $\mathbf{P}$ in Section 3.3.6, which is important because multiplication by all three parts of $\mathbf{C}$ also requires multiplication by $\mathbf{P}$. We discuss a way of accomplishing that in the spirit of [58].[8] All of these speed-ups require small approximations and we assess the effect of those approximations quantitatively. Finally, in Section 3.3.7 we summarize those results and what we can confidently say so far about the accuracy requirements for approximating $\widehat{\mathbf{x}}$ and $\mathbf{P}$ for the purposes of 21 cm power spectrum estimation.

3.3.1 HERA: A Mapmaking Case Study

To test our mapmaking method and our techniques for speeding it up, we need to simulate the visibilities that a real instrument would see. We choose the planned design of the recently commenced Hydrogen Epoch of Reionization Array (HERA) as a particularly timely and relevant case study. HERA will have 331 parabolic dishes, each 14 m in diameter. They will be fixed to point at the zenith with crossed dipole antennas suspended at prime focus. They will be arranged into a maximally

---

[8]The question of preconditioning for rapid conjugate gradient convergence, which was addressed in [58] in the context of estimators based on $\mathbf{x}$ rather than $\widehat{\mathbf{x}}$, is left for future work. That question cannot be answered until the exact form of the $\widehat{\mathbf{x}}$ is chosen. We may choose estimators with a tapering function, such as those suggested by [125] and [126]. We may also choose to project out certain modes from the dirty map, as we discuss in Appendix 3.B.
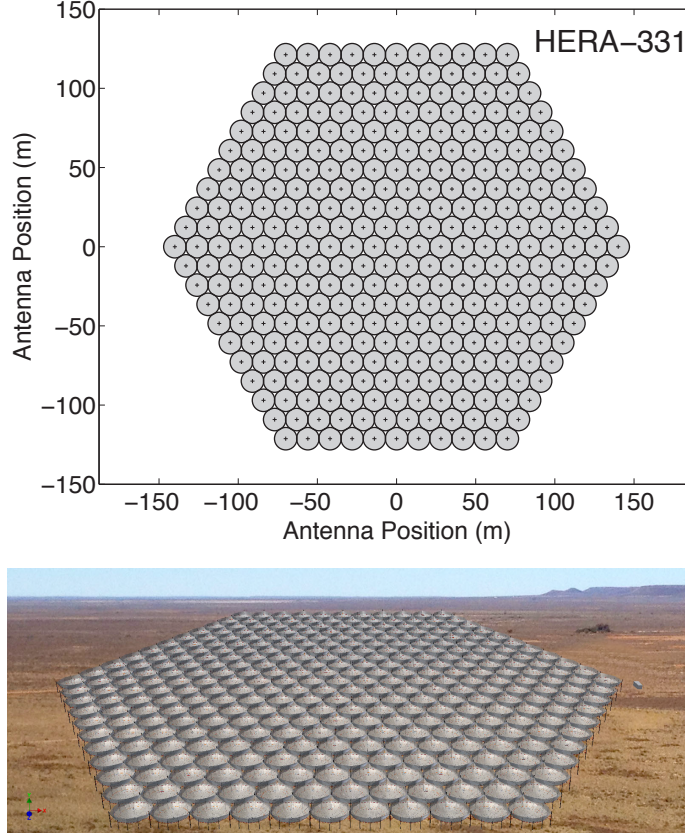
Figure 3-5: We test our method on simulated visibilities from the planned Hydrogen Epoch of Reionization Array (HERA). The array, seeen schematically in the top panel, consists of 331 14 m parabolic dishes, arranged in a close-packed hexagonal configuration. In the bottom panel, we show a rendering of the final array, which will feature more than $0.05\,\mathrm{km}^2$ of collecting area (a standard shipping container, on the right side of the image, is shown for comparison.)

dense hexagonal packing (see Figure 3-5), both to maximize sensitivity to cosmological modes [170, 184] and for ease and precision of calibration [124, 247, 248].[9] In this work, our calculations assume perfect calibration of the instrument and (unless otherwise stated) perfect antenna placement.

HERA also has two advantages that make our algorithms easier to carry out on a relatively small number of computers. First, although it has 331 elements, it only has 630 unique baselines. That is because a highly-redundant array with N baselines

---

[9]Plans for HERA also include outrigger antennas at much greater distances from the hexagonal core to enable low signal-to-noise, high angular resolution imaging. Though they will be useful for making high-resolution maps and modeling astrophysical foregrounds, they do not add significantly to the cosmological sensitivity of the instrument. Since we are focused on maps as a data-compression step between visibilities and power spectra, we ignore them in this analysis.

has $\mathcal{O}(N)$ unique baselines, as opposed to minimally redundant arrays, which have $\mathcal{O}(N^2)$ baselines. That is why the MWA has an order of magnitude more baselines than HERA, even though it has only 128 elements. Second, it has a relatively small primary beam, in contrast to both MWA and PAPER. In this work, we model it fairly accurately as a Gaussian beam with a full width at half maximum of $10°$ at $150\,\mathrm{MHz}$. It should be noted that the method described in this work is independent of the interferometric design. HERA happens to be both a particularly convenient and relevant example.

3.3.2 Testing Mapmaking with a Specific Sky Model

As we find ways to compute mapmaking statistics quickly and accurately, we need to answer a key question: do we understand the relationship between our dirty map $\widehat{\mathbf{x}}$ and the input sky model from which we simulated visibilities? It is not important how much our dirty maps look like the sky itself. We just want to make sure that we keep track of everything the instrument and our mapmaking algorithm has done to the data so we can take it into account properly when start estimating power spectra.

We therefore need an input sky model for two reasons. First, we need to be able to use Equation (3.4) to compute visibilities and thus $\widehat{\mathbf{x}}$. Next, we also want to compute the matrix of point spread functions $\mathbf{P}$ corresponding to the same set of observations and multiply it by our true sky model $\mathbf{x}$. The error metric we use therefore is

$$\varepsilon = \frac{|\widehat{\mathbf{x}}_{\mathrm{exact}} - \widehat{\mathbf{x}}_{\mathrm{approx}}|}{|\widehat{\mathbf{x}}_{\mathrm{exact}}|}. \tag{3.29}$$

To be clear, this does not measure the difference between our dirty map and the true sky. It is merely a measure of the discrepancy between what the instrument and our mapmaking routine did to the sky in order to form the dirty map ($\widehat{\mathbf{x}}_{\mathrm{exact}}$) and what we think we know about those effects ($\widehat{\mathbf{x}}_{\mathrm{approx}}$) when we write down $\boldsymbol{\mu}$ and $\mathbf{C}$.

One advantage to this metric is that it is often relatively easy to calculate $\widehat{\mathbf{x}}_{\mathrm{exact}}$, at least up to $\mathbf{D}$ which we can factor out of the numerator of Equation (3.29), compared to calculating $\mathbf{P}$. That is because calculating $\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{y}$ is as computationally difficult

as calculating a single row of $\mathbf{P}$. In the following sections, we will be examining ways of computing $\mathbf{P}$ faster. Sometimes (e.g. in Sections 3.3.4 and 3.3.6) that means an approximate $\mathbf{P}$ but an exact $\widehat{\mathbf{x}}$, in which case $\widehat{\mathbf{x}}_{\mathrm{approx}} = \mathbf{P}_{\mathrm{approx}}\mathbf{x}$. Other times (e.g. in Section 3.3.5) that means a method for computing $\widehat{\mathbf{x}}$ that also makes $\mathbf{P}$ easier to compute. In that case, Equation (3.29) compares the approximate method for computing $\widehat{\mathbf{x}}$ with the exact one.

We have chosen a sky model with two components: 1) bright point sources and 2) diffuse emission from our Galaxy and other dim, confusion-limited galaxies. Since each frequency is measured and analyzed independently (meaning that $\mathbf{A}$ is sparse and can be written compactly in blocks), we will perform all the simulations at a representative frequency of 150 MHz. While the simulations properly weight visibilities based on how many times each unique baseline was measured, we do not include any noise in our calculation of the quantities in Equation (3.29). We also assume that all baselines at a given frequency have the same noise properties, though that assumption can be straightforwardly relaxed.

3.3.2.1 Point Sources

Our sky model includes bright point sources above 1 Jy with specified positions, fluxes, and spectral indices. These are taken from the MWA Commissioning Survey Catalog [97], which is complete to below 1 Jy for a large fraction of the sky. The included spectral indices are used to extrapolate their fluxes at 150 MHz down from the survey frequency of 180 MHz. For the calculation of visibilities using Equation (3.4), they are treated as true point sources with Dirac delta function spatial extent. In Figure 3-6, we show a representative sample of those point sources and what they look like in the dirty map, $\widehat{\mathbf{x}}$.

The sky model for point sources is completely independent of our pixelization. Since we know the location of all the point sources, we can think of $\mathbf{x}$ as having a discretized component covering the whole sky in pixels—which we will use for analyzing diffuse emission—and a set of Dirac delta function fluxes at the positions of the point sources. The sky model for point sources is completely independent of
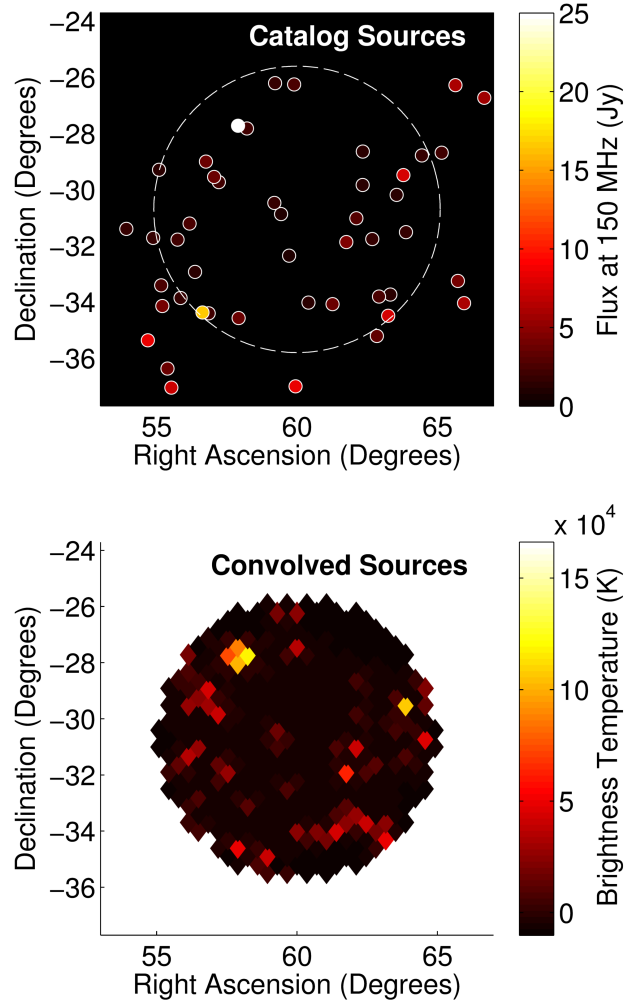
Figure 3-6: To test our mapmaking method and our approximate techniques for making it much faster, we need a fiducial sky model. One component of that model is bright point sources, which are taken from the MWA Commisioning Survey Catalog [97]. In the top panel, we show the spatial distribution and intrinsic flux of all point sources whose primary-beam-weighted fluxes are above 1 Jy. In the bottom panel, we show $\widehat{\mathbf{x}} = \mathbf{Px}$, the PSF-convolved and discretized dirty map with HEALPix $N_{\text{side}} = 128$. Since the point spread functions are computed at the locations corresponding to each point source, the bottom panel is exact.

our pixelization. This is completely compatible with the definition of our pixelization in Section 3.2.1, it is just that some pixels have finite area and some have infinitesimal area. It is the pixels with finite volume that we care about for 21 cm power spectrum estimation, but the infinitesimal "pixels" matter for foreground subtraction. Likewise, $\mathbf{P}$ has two blocks: one that maps pixels on the true sky to pixels on the dirty map and one that maps points on the true sky to pixels on the dirty map.

3.3.2.2 Diffuse Emission

In the case of point sources, we might hope to use precise locations on the sky to refine our models of $\boldsymbol{\mu}$ and $\mathbf{C}$ and do a better job of separating foregrounds from the 21 cm signal. That is simply not possible with diffuse synchrotron emission from our Galaxy or with the confusion-limited emission from relatively dim radio galaxies. Fundamentally, our best guess at that emission and its statistics will have to be discretized and pixelized. Uncertainty about how many confusion-limited point sources appear in a single pixel introduces shot noise, which can be modeled [120, 58].

In this work, we are interested in errors caused by assumptions and approximations in our mapmaking routine whose effects are not taken into account when estimating power spectra. In order to write down a vector $\mathbf{x}$ that we can use to compute $\widehat{\mathbf{x}}$ and thus ε with Equation (3.29), we can either treat the emission as constant in the pixel or we can treat the emission as a "point source" at the center of each pixel. For computational simplicity, we choose the latter. With relatively small pixels, there is no practical difference between the two. Since we are concerned about translating our models for foreground residuals in the true sky into models in the dirty map, the pixelization here is not an approximation so much as a consequence of the discretized models for foreground residuals we need for power spectrum estimation. It is possible to construct $\mathbf{P}$ to have different angular resolutions of $\mathbf{x}$ and $\widehat{\mathbf{x}}$, if one would like to incorporate a high-resolution diffuse foreground covariance model. The more information we can incorporate about the foregrounds, the smaller our uncertainties get and the better foreground subtraction works.

We use the popular HEALPix software package [79] for discretizing the celestial

169

sphere into regularly spaced, equal-area pixels. As a model for the emission itself, we use the Global Sky Model of de Oliveira-Costa et al. [54] (see Figure 3-7). The precise model we choose for this work matters only insofar as it is relatively realistic and representative of the true sky. That said, building good foreground models is an important ongoing endeavor relevant to power spectrum estimation and foreground subtraction [234, 182, 104, 244, 109].

3.3.3 Computational Challenges of Mapmaking

We already alluded to the fact that we need to investigate various simplifications and approximations to make the calculation of $\hat{\mathbf{x}}$ and $\mathbf{P}$ tractable. Let us take the time to see exactly where the problem lies.

Consider the matrix $\mathbf{A}$ where $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$. $\mathbf{A}$ maps a discretized sky into time-ordered data. If we want to slightly over-resolve the sky with HERA, we might choose a HEALPix map with $N_{\mathrm{side}} = 256$, which gives an angular resolution of about $0.2°$. That is almost 10^6 pixels at each of about 1000 different frequencies (assuming $100\,\mathrm{kHz}$ resolution and $100\,\mathrm{MHz}$ of simultaneous bandwidth). If we measure all our visibilities every two seconds for 1000 total hours at all 1000 frequencies, that is 10^{14} visibilities, so naively, $\mathbf{A}$ is a $10^{14} \times 10^9$ matrix. That is a problem.

Of course, there are many standard simplifications. Each frequency is treated completely independently during mapmaking, so we can treat $\mathbf{A}$ as either block diagonal or as a family of 1000 much smaller matrices, $\mathbf{A}(f)$. Redundant baselines measure the same sky, so their visibilities can be combined together, reducing both instrumental noise and the number of visibilities by a factor of almost 100 in the case of HERA. Getting 1000 hours of nighttime observation takes about 100 days, so we can LST-bin, reducing both noise variance and data volume by another two orders of magnitude. Since each time-step is independent of all others, we can further break $\mathbf{A}$ into about 10,000 pieces for each integration.

We still have 10^7 different $\mathbf{A}$ matrices, each $10^3 \times 10^6$. This size is challenging but acceptable for either simulating visibilities or calculating $\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{y}$. However, it is simply too big for the calculation of $\mathbf{P}$, which would require the computationally
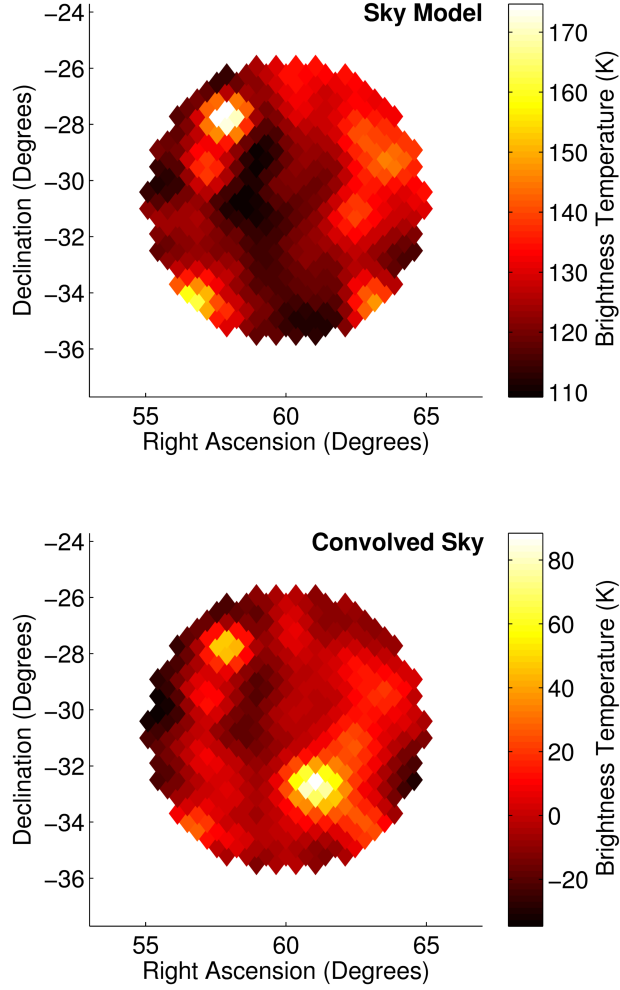
Figure 3-7: The sky model we use to evaluate our mapmaking algorithm and the accuracy of the approximations we make also includes diffuse emission from our Galaxy and faint radio galaxies. For our model of diffuse emission, we use the Global Sky Model of [54]. In the top panel, we show a small part of our model for the true diffuse emission. Since we are not trying to model fine spatial information or the precise locations of point sources with our diffuse models, we pixelize the emission identically to the pixelization of our dirty map. In the bottom panel, we show that dirty map. It looks fairly different from the true sky, largely because of the appearance of a side lobe from a bright object outside the field. This occurs because the **P** maps a very large region of the sky to a small one shown here. The effects of faceting and side lobes will be explored further in Section 3.3.4.

infeasible task of multiplying together two matrices of this size 10^7 times, each multiplication taking roughly 10^{15} operations. In the following sections, we will look at ways of reducing the number of $\mathbf{A}(f)$ matrices and making each $\mathbf{A}(f)$ smaller, especially during the calculation of $\mathbf{P}$.

3.3.4 Faceting and First Mapmaking Results

The matrix of point spread functions $\mathbf{P}$ is defined by the relation $\langle \widehat{\mathbf{x}} \rangle = \mathbf{P}\mathbf{x}$. It can be thought of as a transformation from one pixelized real space—that of the true sky—to another—that of the dirty map. For even a modest angular resolution, that is an enormous matrix. Do we really need to know the relationship between every pixel in the sky and every pixel in the dirty map?

3.3.4.1 Why We Facet

Breaking up the field of view into a number of smaller facets is a standard technique in radio astronomy, especially when one wants to minimize the effects of noncoplanar baselines [47]. For purposes of 21 cm cosmology, there are two good reasons to consider relatively small regions of the sky one at a time. The first is HERA's observing strategy. Because it statically points at the zenith, HERA scans a fixed stripe in declination about 10° degrees wide. It seems reasonable that we can analyze parts of the stripe independently, making maps and computing power spectra for each small facet. In Figure 3-8, we show an example of what that faceting might look like.

The only significant disadvantage to faceting is that we lose the ability to measure modes in the power spectrum with wavelengths perpendicular to the line of sight that are larger than the facet. Doing so properly and with precisely quantified error properties would require calculating covariance between facets, which is effectively the same as not faceting at all. This is not such a great hardship. Due to the survey geometry, only the long modes oriented along the HERA stripe could have been measured at all. They are longer than the shortest baseline, meaning that they can only be sampled after considerable sky rotation. The same $|\mathbf{k}|$ modes can be also
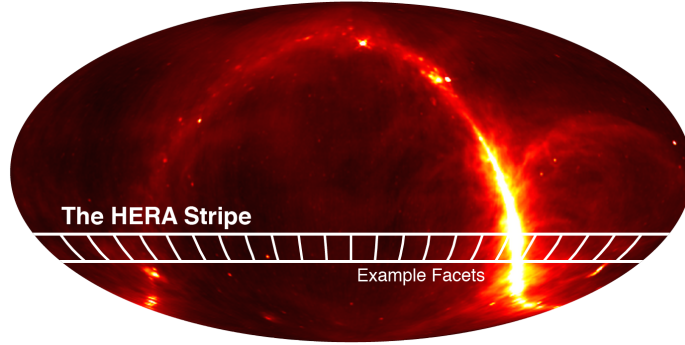
Figure 3-8: The faceted approach we use to speed up optimal mapmaking and power spectrum estimation will be especially useful for HERA because it is limited to only observe an approximately $10°$ stripe of constant declination, centered on the array's latitude of approximately $-30.7°$. It is fairly natural to split up the observation into roughly $10° \times 10°$ facets, each analyzed separately. This makes **P** much easier to compute and lets us use the flat-sky approxmation, a requirement for implementing the power spectrum methods of Dillon et al. [58]. Very little cosmological information is lost in this process; only the longest spatial modes are thrown out and they should be dominated by galactic emission.

be accessed along the line of sight, except those at very low spectral wave-numbers, which are bound to be foreground dominated.

The other major upside to faceting is that, if we want to use the fast power spectrum techniques developed in [58], we need to take our maps and chop them up into facets anyway. That is because any fast algorithm that takes advantage of the fast Fourier transform (e.g. that in [58]) and translational invariance relies on rectilinear data cubes, which is only an accurate approximation for small fields where the flat-sky approximation holds. Happily, that rough size is also about $10°$. For other instruments, the choice of facet size is less obvious and depends on the computational demands of both mapmaking and power spectrum estimation. Bigger facets preserve more information, but they can be more computationally expensive than they are cosmologically useful. The exact right choice for other interferometers is a matter for future work.

3.3.4.2 Faceted Mapmaking Method And Results

So, instead of using $\mathbf{DA}^{\dagger}\mathbf{N}^{-1}\mathbf{y}$ to calculate $\widehat{\mathbf{x}}$, we instead redefine $\widehat{\mathbf{x}}$ using

$$\widehat{\mathbf{x}} = \mathbf{DK}_{\text{facet}}\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{y}, \tag{3.30}$$

where $\mathbf{K}_{\text{facet}}$ maps the full sky to a small portion of the sky, thus making $\mathbf{P}$ asymmetric. Doing this for every facet basically amounts to only mapping the parts of the sky that are ever near the center of the primary beam. This provides a computational simplification by a factor of $4\pi/(\Omega_{\text{facet}}N_{\text{facets}})$, which for HERA is about an or order of magnitude. An instrument that can see the whole sky would see no computational benefit just from breaking the sky in facets.

The real computationally limiting step is the calculation of $\mathbf{P}$. Since we are only interested in the dirty map of a facet, we care only about source flux that could have contributed to that dirty map. That means that we can truncate each point spread function some distance from the facet center. Flux outside that truncation radius is assumed not to contribute significantly. In other words,

$$\mathbf{P} = \mathbf{DK}_{\text{facet}}\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{AK}_{\text{PSF}}^{\mathsf{T}} \tag{3.31}$$

where $\mathbf{K}_{\text{PSF}}$ is the same as $\mathbf{K}_{\text{facet}}$ except that it cuts off at some larger radius than the facet size. We get to choose exactly what radius we want to assume that no outside flux contributes to the facet. This is a completely tunable approximation and it becomes exact in the limit that that radius encompasses the whole sky.

Therefore, instead of mapping the whole sky to the whole sky, the matrix of point spread functions now maps some moderate portion of the sky to a somewhat smaller part of the sky. Since $\mathbf{N}$ is diagonal, both the time it takes to calculate $\mathbf{P}$ and the memory it takes to store it are reduced by very large factor. If the truncation region is 4 times the $10°$ facet size, for example, then that savings is a factor of about 10^4.

This new definition of $\widehat{\mathbf{x}}$ means that $\mathbf{D}$ is now a much, much smaller matrix—it has only as many elements as there are pixels in the facet. And since we are only

interested in the correlation between pixels in the map, the noise covariance is now

$$\mathbf{C}^N = \mathbf{P}\mathbf{K}_{\text{facet}}^{\mathsf{T}}\mathbf{D}^{\mathsf{T}}, \tag{3.32}$$

which is much smaller and still quite simple.

We illustrate the effect of the PSF truncation radius in Figure 3-9, showing the large impact that increasing the truncation radius has on our calculations of $\hat{\mathbf{x}}_{\text{approx}} = \mathbf{P}_{\text{approx}}\mathbf{x}$ and therefore of ε. We find that once the PSF includes both the central peak of the synthesized beam and the first major side lobes, the convergence of $\hat{\mathbf{x}}_{\text{approx}}$ to $\hat{\mathbf{x}}_{\text{exact}}$ is very quick.

We further tested the expected convergence of the algorithm for a fixed facet size and variable $\mathbf{K}_{\text{PSF}}$ using the sky model from Section 3.3.2. Our results, which we show in Figure 3-10, again demonstrate that the PSF truncation radius does not need to be much larger than the facet, if the facet is comparable in size to the primary beam. The exact level of error introduced by faceting will, in general, depend upon the compactness of both the primary and synthesized beams. The approximation that the point spread function is Gaussian might make the plotted relative error a bit optimistic, though the side lobes in the real HERA primary beam are quite small.

In summary, faceting allows us to decrease the time it takes to calculate the $\mathbf{P}$ and the memory required to store it by a factor of $(4\pi)^2/(\Omega_{\text{facet}}\Omega_{\text{PSF}})$, where Ω_{PSF} is the angular size of the region left by $\mathbf{K}_{\text{PSF}}$. In the case of HERA, that works out to about 10,000 times faster and smaller.

3.3.4.3   Mitigating Nonredundancy

Making maps in facets also has one extra advantage useful in addressing a common complication presented by real-world arrays. If we assume in our analysis that every baseline of a given designed separation actually has that separation, we will be ignoring errors that can be a decent fraction of a wavelength. And though HERA is a zenith-pointed array for which noncoplanar effects are small, they are not zero and can be quite large for other instruments like the MWA. Noncoplanarity creates
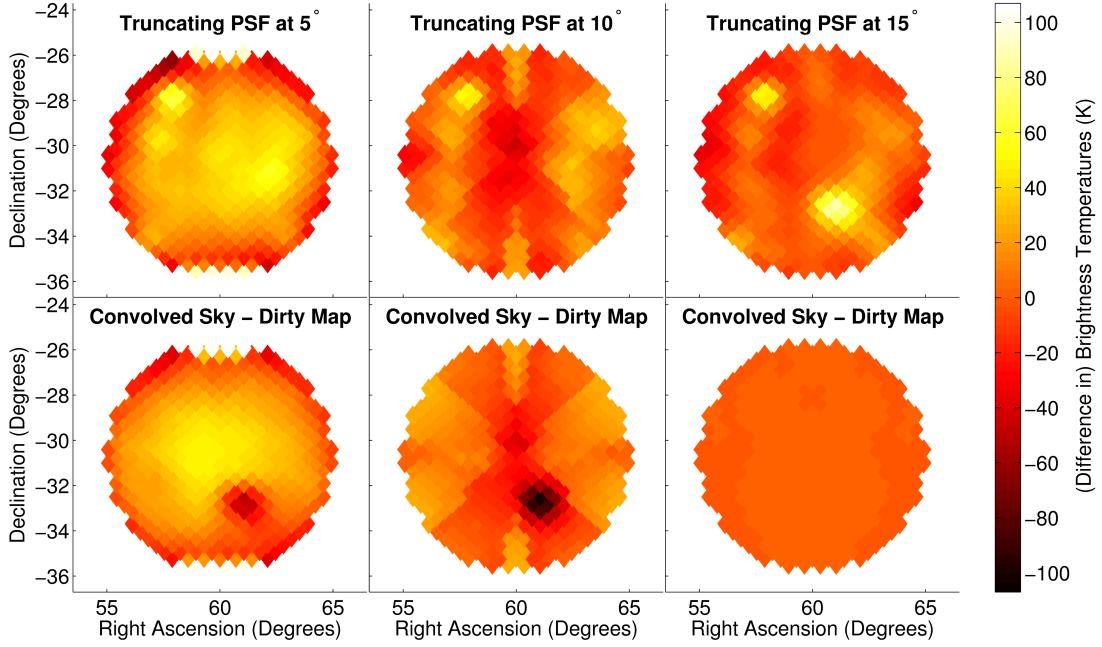
Figure 3-9: In order to accurately reproduce dirty maps, we must include in our $\mathbf{P}$ matrix the effect flux from outside the facet that appears in the side lobes of off-facet sources. Here we demonstrate that effect by looking at how the approximate PSF-convolved sky, $\mathbf{P}_{\mathrm{approx}}\mathbf{x}$, evolves as we expand the distance from the center of the facet at which the point spread function is approximated to not contribute. In the top row, we plot $\mathbf{P}_{\mathrm{approx}}\mathbf{x}$ while on the bottom row we plot $\mathbf{P}_{\mathrm{approx}}\mathbf{x}-\widehat{\mathbf{x}}_{\mathrm{exact}}$. ($\mathbf{P}_{\mathrm{exact}}\mathbf{x}=\widehat{\mathbf{x}}_{\mathrm{exact}}$ is shown in the bottom panel of Figure 3-7.) Since the visibilites that go into computing $\widehat{\mathbf{x}}$ derive from a full-sky calculation, side lobes are automatically included. The bright spot we see on the top right panel, which appears as a dark spot on the bottom left and bottom middle panels, is a prominent side lobe from a very bright source outside the facet, but within $15°$ of the facet center. This explains what we saw in Figure 3-7 and the dramatic improvement in the error we see in the right-hand panels.
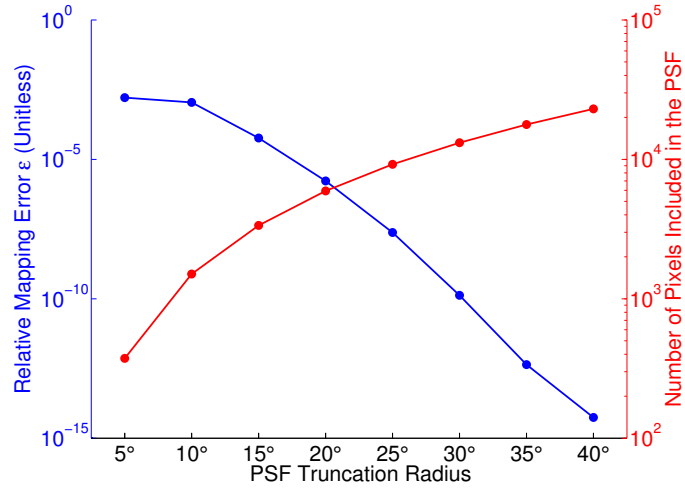
Figure 3-10: The error introduced by the approximation that the PSF can be truncated past a certain distance from the facet center gets very small very quickly. Here we show how both that error, which we define in Equation (3.29), and the number of pixels in each point spread function depend on the truncation radius. The number of pixels, and thus the computational difficulty of computing the matrix of point spread functions, $\mathbf{P}$, scales as the truncation radius squared—there are simply more pixel values to calculate. In general, the approximation works because the point spread functions are relatively compact. HERA's design is especially helpful here with its dense grid of baselines and its relatively small primary beam. Other arrays may need larger truncation radii to acheive the same accuracy.

nonredundancy.

However, as long as we know precise positions of all of our antennas (which is far easier than making the array perfectly redundant) we can use the fact that we are only mapping a single facet at a time to reduce those phase errors near the center of our map. We can think of each baseline corresponding to some unique baseline $\mathbf{b}$ as

$$\mathbf{b}_m = \mathbf{b} + \Delta\mathbf{b}_m, \tag{3.33}$$

where the residuals are caused by inexact antenna placement. That means that Equation (3.7) becomes

$$V(\mathbf{b}_m, \nu_n) \approx \sum_k \Delta\Omega \frac{2k_B\nu_n^2}{c^2} x_k(\nu_n) B(\hat{\mathbf{r}}_k, \nu_n) \times$$
$$\exp\left[-2\pi i \frac{\nu_n}{c} (\mathbf{b} + \Delta\mathbf{b}_m) \cdot \hat{\mathbf{r}}_k\right]. \tag{3.34}$$

We need the right-hant side of this equation to be the same for all $\mathbf{b}_m$ corresponding to the unique baseline $\mathbf{b}$, otherwise we lose the redundancy bonus we discussed in Section 3.3.3.

We can achieve this approximately for small $\Delta\mathbf{b}_m$ because our facets are relatively small. Let us define $\Delta\hat{\mathbf{r}}_k \equiv \hat{\mathbf{r}}_k - \hat{\mathbf{r}}_0$ where $\hat{\mathbf{r}}_0$ points to the center of the facet and $\Delta\hat{\mathbf{r}}_k$ is generally not a unit vector. We can expand the exponent of Equation (3.34) as

$$(\mathbf{b} + \Delta\mathbf{b}_m) \cdot (\hat{\mathbf{r}}_0 + \Delta\hat{\mathbf{r}}_k)$$
$$= \mathbf{b} \cdot \hat{\mathbf{r}}_0 + \mathbf{b} \cdot \Delta\hat{\mathbf{r}}_k + \Delta\mathbf{b}_m \cdot \hat{\mathbf{r}}_0 + \Delta\mathbf{b}_m \cdot \Delta\hat{\mathbf{r}}_k. \tag{3.35}$$

The first two terms in the expansion are $\mathbf{b} \cdot \hat{\mathbf{r}}_k$ and normally appear in $\mathbf{A}$. The last term, which second order in this expansion, is approximated to be zero. Even if $\mathbf{b} \cdot \hat{\mathbf{r}}_0$ is small, the last term is in general much smaller than the second term. We can, however, correct for the middle term by multiplying both sides of Equation (3.34) by

a constant phase factor, since

$$V(\mathbf{b}, \nu_n) \approx \exp\left[2\pi i \frac{\nu_n}{c} \Delta\mathbf{b}_m \cdot \hat{\mathbf{r}}_0\right] V(\mathbf{b}_m, \nu_n). \tag{3.36}$$

As was our goal, the $\mathbf{P}$ matrix that results from taking the above equation to be exactly true is the same as if we had not had any antenna placement errors or non-coplanarity. Rephasing lets us mitigate the effect of known errors without having to calculate a vastly more complicated $\mathbf{P}$, which treats all baselines completely independently, even if they are supposed to be redundant.

Effectively, our approximate correction cancels out the phase error at the exact center of the facet and thus minimizes its effect throughout the facet. For example, for 10° facets at 150 MHz, a 4 cm antenna placement error (roughly the level seen in [248]) leaves only a 0.63° phase error in the visibility after rephasing. The error might be a bit worse when calculating the parts of $\mathbf{P}$ near the truncation radius. For very large fields, as [47] addressed, this becomes a bigger problem and we may need to break each set of baselines that was supposed to be redundant into a few groups, each closer to exactly redundant, and treat each group separately. The exact effect on the accuracy of the dirty maps from this small correction is left to future work when the exact antenna placement of HERA or a similar array is known.

3.3.5 Grouping Visibilities into Snapshots

Standard interferometric mapmaking techniques accumulate visibilities in the uv-plane via sky rotation and thereby combine minutes or even hours of visibilities together [196, 49, 48, 21, 35, 149, 19, 177, 201, 116, 203, 162, 224]. We would like to find a way of reducing the number of rows in $\mathbf{A}$ for the purpose of calculating $\mathbf{P}$ by grouping integrations into "snapshots" that are each analyzed as a single timestep when we calculate $\mathbf{P}$. How can we average together multiple visibilities over a range of times while approximating the $\mathbf{P}$ as having been calculated at only the middle timestep of each snapshot?

Once again, we can use our freedom to rephase both the visibilities and the $\mathbf{A}$

matrix as we did in Section 3.3.4.3. The idea is to try to remove, as much as possible, the effect of sky rotation from the visibilities. Consider again Equation (3.4), now with explicit time dependence:

$$V(\mathbf{b}, \nu, t) = \int B(\hat{\mathbf{r}}, \nu) I(\hat{\mathbf{r}}, \nu, t) \times$$
$$\exp\left[-2\pi i \frac{\nu}{c} \mathbf{b} \cdot \hat{\mathbf{r}}\right] d\Omega. \tag{3.37}$$

While the sky rotates, the primary beam is fixed relative to the ground.

By contrast, let us consider a new reference frame with angle vector $\hat{\mathbf{r}}'$, which rotates with the sky:

$$V(\mathbf{b}, \nu, t) = \int B(\hat{\mathbf{r}}', \nu, t) I(\hat{\mathbf{r}}', \nu) \times$$
$$\exp\left[-2\pi i \frac{\nu}{c} \mathbf{b}(t) \cdot \hat{\mathbf{r}}'\right] d\Omega'. \tag{3.38}$$

Now the beam and the baseline vector have picked up an explicit time dependence while the sky has lost its time dependence. Let us assume that the primary beam is varying very slowly spatially—generally a good assumption since the primary beam is much larger than the spatial scales probed by most baselines.

Let us think of $V(\mathbf{b}, \nu, t)$ as the visibility measured for the middle integration of a snapshot. A visibility measured a bit later during that snapshot would look like

$$V(\mathbf{b}, \nu, t + \Delta t) \approx \int d\Omega' B(\hat{\mathbf{r}}', \nu, t) I(\hat{\mathbf{r}}', \nu) \times$$
$$\exp\left[-2\pi i \frac{\nu}{c} (\mathbf{b}(t) + \Delta \mathbf{b}) \cdot \hat{\mathbf{r}}'\right], \tag{3.39}$$

where $\Delta \mathbf{b}$ is the difference between $\mathbf{b}(t + \Delta t)$ and $\mathbf{b}(t)$ in the primed coordinate system. The dot product is basis independent, so

$$(\mathbf{b}(t) + \Delta \mathbf{b}) \cdot \hat{\mathbf{r}}' = \mathbf{b} \cdot (\hat{\mathbf{r}} + \Delta \hat{\mathbf{r}}(\hat{\mathbf{r}})), \tag{3.40}$$

where the right-hand side is back in the frame that is stationary relative to the Earth.

$\Delta\hat{\mathbf{r}}(\hat{\mathbf{r}})$, which is not a unit vector, is the amount of sky rotation between times t and $t + \Delta t$. It is approximately constant across the facet for fairly short snapshots and moderately sized facets, meaning that we can pull it out of the integral. We can therefore undo much of the effect of sky rotation using the approximation that

$$V(\mathbf{b}, \nu, t + \Delta t) \approx e^{i\Delta\phi} V(\mathbf{b}, \nu, t) \tag{3.41}$$

where

$$\Delta\phi \equiv -2\pi \frac{\nu}{c} \mathbf{b} \cdot (\hat{\mathbf{r}}_0(t + \Delta t) - \hat{\mathbf{r}}_0(t)) \tag{3.42}$$

and where again, $\hat{\mathbf{r}}_0(t)$ points to the facet center.

We can therefore add together many visibilities taken at different times and approximately treat them as if there were all taken at the middle integration in the snapshot by rephasing them. This is very similar to the "fringe-stopping" technique from traditional radio astronomy, which seeks to counteract the effect of the rotation of the earth at the location of a source [216]. As we saw in Section 3.3.4.3, the effect of rephasing visibilities cancels out in $\mathbf{P}$, since the extra term in $\mathbf{A}$ gets canceled out in $\mathbf{A}^\dagger$. That is why we only have to perform the calculation of $\mathbf{P}$ once per snapshot rather than once per integration. We show in Figure 3-11 a marked improvement, especially in the case of long snapshots, between naively adding together visibilities as if the sky were not rotating overhead and adding together rephased visibilities.

In Figure 3-12 we show quantitatively how the error increases as snapshots get longer. Here we care how these approximate dirty maps compare to the exact dirty maps made when each $10\,\mathrm{s}$ integration is treated completely separately. We also found it important to rephase the visibilities to the exact middle of the snapshot, which creates a first-order cancellation that removes some of the error associated with this approximation.

Based on the results we show in Figure 3-12, it is likely that we can cut another one to two orders of magnitude off the total number of operations we need to perform to calculate $\mathbf{P}$, making that calculation considerably easier. For a given accuracy goal, it is also possible to make the calculation of $\mathbf{P}$ even simpler by forming snapshots
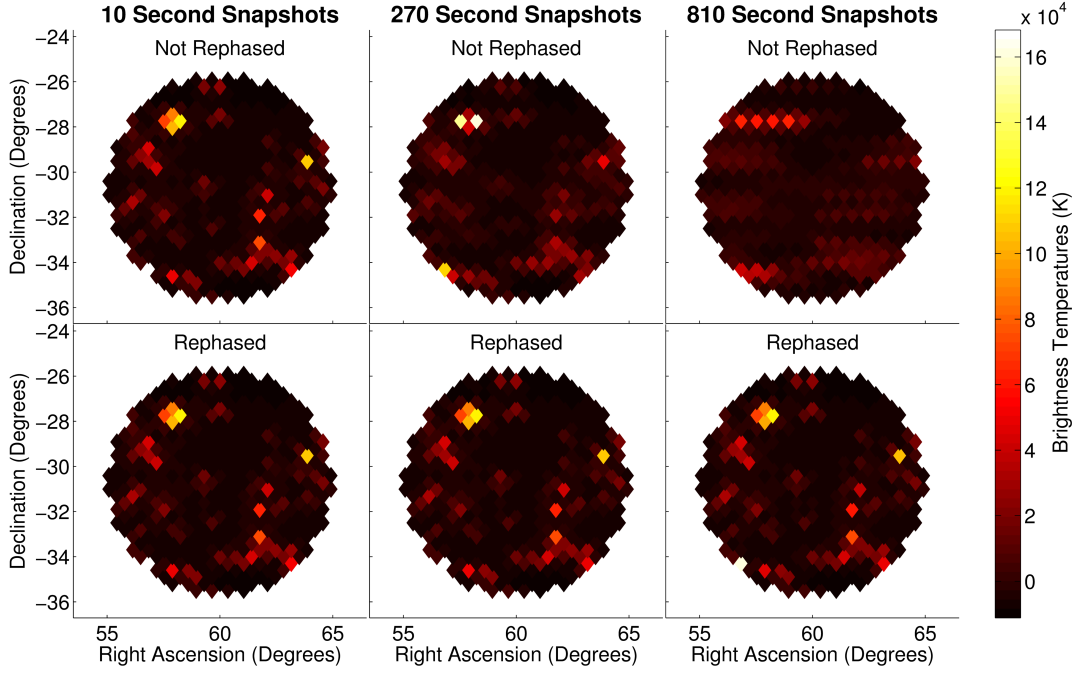
Figure 3-11: One way to make the calculation of the matrix of point spread functions, **P**, faster is to combine many consecutive integrations together into snapshots. When we compute **P**, we effectively assume that all the associated visibilities we have grouped into one snapshot were taken exactly at the snapshot's middle time. Usually, this is a poor approximation. As we can see from the top row, where we have simply added together 10 second integrations to snapshots of increasing length, we are effectively spreading out sources in right ascension as the sky rotates overhead. However, if we use our freedom to rephase visibilities individually, we can dramatically reduce the error associated with forming snapshots. For example, the bottom right panel only exhibits error on the order of a few percent compared to the exact single-integration dirty maps in the left-hand panels. The result is related to the traditional radio astronomy technique of "fringe stopping."
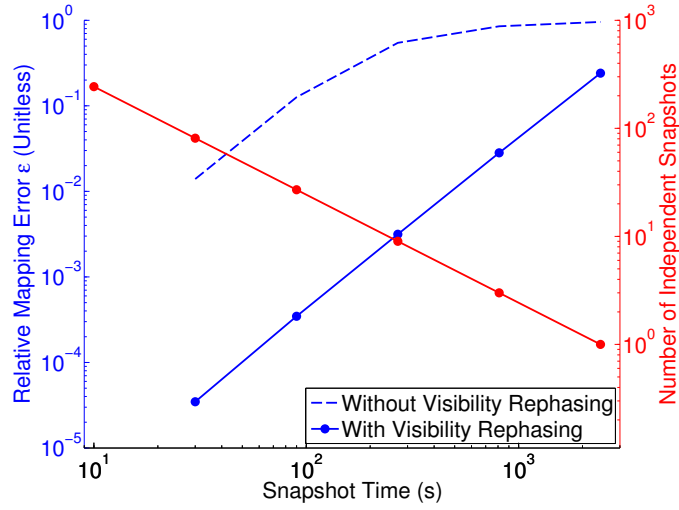
Figure 3-12: The error introduced by approximating the observation as having taken place at at only a few discrete times, many seconds or minutes apart, can be mitigated by appropriately rephasing visibilities before combining them. Here we show quantitatively how the length of snapshots—all multiples of the 10 second integration time used in our simulation—introduces small errors. We calculate the relative error ε between dirty maps calculated with a given integration time and those calculated exactly using only one integration per snapshot. We also show how the computational difficulty of calculating $\mathbf{P}$ is affected, since it scales linearly with the number of independent snapshots considered.

with different durations for baselines of different lengths, keeping $\Delta\phi$ small.

3.3.6 PSF Fitting

Now that we have found accurate and well-understood approximations that make computing $\mathbf{P}$ computationally feasible, we need to worry about multiplying a vector by $\mathbf{P}$. This is a necessary step in any power spectrum estimation scheme adapted from [58], since $\mathbf{P}$ appears in Equations (3.23), (3.24), (3.26), and (3.28). In general, the number of operations in this calculation scales with the number of pixels in the facet, the number of pixels in the PSF, and the number of frequency channels, i.e. as $\mathcal{O}(N_{\mathrm{facet}}N_{\mathrm{PSF}}N_f)$. This is slower than we would like, so we will endeavor to show how it can be sped up.

If the point spread function were constant across the field—if it looked the same in the top and bottom rows of Figure 3-3—then the solution would be simple. We could calculate only one PSF and then use it to fill out all of $\mathbf{P}$. Then, if we approximate HEALPix pixelization as a regular grid—which is true in the flat-sky approximation—we can write $\mathbf{P}$ using Toeplitz matrices. A Toeplitz or "constant-diagonal" matrix represents a translationally invariant relationship.[10] A Toeplitz matrix $\mathbf{T}$ has the property that each element only depends on its distance from the diagonal of the matrix, or in other words that

$$T_{ii'} = t_{i-i'}. \tag{3.43}$$

We can imagine that, if any part of the PSF can be fully represented by its displacement from the facet center, then we can write $\mathbf{P}$ for each frequency and facet as a tensor product of two matrices, each describing translational invariance along one of the two principal axes of the HEALPix grid.[11] If we index along those axes with i and j in the dirty map and i' and j' in the true sky, then for a single frequency the

---

[10]Toeplitz matrices have a number of nice properties, including the fact that an $N \times N$ Toeplitz matrix can be multiplied by a vector in $\mathcal{O}(N \log N)$ operations. This is because the translational invariance lets us use the fast Fourier transform. See [80] for a review of these matrices and their properties or [58] for a previous application to 21 cm cosmology of the same relevant properties.

[11]We define the axes by taking the center pixel and computing the linearly independent vector directions towards the nearest two pixels. It is not a problem that these two directions are not orthogonal—the FFT can be performed along nonorthogonal directions, as pointed out by [211].

matrix of point spread functions can be written as

$$P_{ii'jj'} = t_{i-i'}s_{j-j'} \tag{3.44}$$

or as

$$\mathbf{P} = \mathbf{T} \otimes \mathbf{S} \tag{3.45}$$

where $\mathbf{T}$ and $\mathbf{S}$ are Toeplitz matrices.

And yet we can easily see from Figure 3-3 that point spread functions do not respect translational invariance. In the bottom row where the PSFs are displaced from the center of the facet, the side lobes nearer the edge of the primary beam are downweighted relative to those nearer the center. This is a consequence of optimal mapmaking, which downweights the contribution from regions of the sky that the telescope is less sensitive to. However, we expect that the physical effects that lead to a translationally varying PSF, like the primary beam and the projected array geometry, should change smoothly over the field. So while the PSF is translationally varying, perhaps its translational variation can be modeled with a small number of parameters.

If we calculate $\mathbf{P}$, the matrix of point spread functions that maps every pixel in some extended facet to every pixel on the facet of interest, we can model this translational variation by reorganizing $\mathbf{P}$. We have chosen our normalization $\mathbf{D}$ so that the specific point spread function mapping the sky onto a given pixel has a value of 1 at the center pixel of its main lobe. But what about all the pixels displaced exactly pixel northeast from the center of the main lobe in all the PSFs? Or ten pixels?

We expect these all to be similar, but also to vary slowly over the facet—though exactly how is not obvious *a priori.* In the right-hand panel of Figure 3-13 we plot the points on the PSFs displaced exactly 15 pixels along one of the two principal axes from the centers of their main lobes (illustrated by the left-hand panel). The x and y axes of the plot tell us which pixel a given PSF is centered on. As we expected, the variation over the facet is very smooth and is well approximated by a low-order

polynomial. If we had instead plotted a displacement of 0, the right-hand panel would have been a perfectly flat plan of all ones because of the definition of $\mathbf{D}$.

How can we take advantage of the sparsity of information needed to describe $\mathbf{P}$ to write it as the sum of matrices that can be quickly multiplied by a vector? Let us first consider the simpler, 1D case. Instead of the translational invariance that leads to matrices of the form in Equation (3.43) where the main diagonal and all parallel off diagonals are constant, instead we model them all as polynomials:

$$P_{ii'}^{1D} = \sum_n t_{n,i-i'}(i+i')^n. \tag{3.46}$$

This is a polynomial expansion in $(i + i')$, the distance along a diagonal, with coefficients $t_{n,i-i'}$ that make up a Toeplitz matrix. Again, primed indices tell us where on the true sky and unprimed indices tell us where in the faceted dirty map. The polynomial fit coefficients are a function of specific displacement of the main lobe of the PSF, hence the index $i - i'$. However, to fit all PSF values for the same displacement, we need to multiply those coefficients by the displacement from the center of the facet to the correct polynomial power. Our hope is that we can approximate $\mathbf{P}$ with a relatively low-order polynomial.

Expanding this out and cutting off the series after the second order in n, we get that

$$\mathbf{P}^{1D} \approx \mathbf{T}_0 + \mathbf{J}\mathbf{T}_1 + \mathbf{T}_1\mathbf{J} + \mathbf{J}^2\mathbf{T}_2 + 2\mathbf{J}\mathbf{T}_2\mathbf{J} + \mathbf{T}_2\mathbf{J}^2 \tag{3.47}$$

where each $\mathbf{T}_n$ is a Toeplitz matrix and $\mathbf{J}$ is a diagonal matrix with integer indices centered on zero as its entries:

$$\mathbf{J} \equiv \mathrm{diag}\left(..., -4, -3, -2, -1, 0, 1, 2, 3, 4, ...\right). \tag{3.48}$$

Terms in the expansion that involve $(i')^n$ look like $\mathbf{J}^n$ to the right of $\mathbf{T}_n$, since they index into a vector multiplied by $\mathbf{P}^{1D}$ on the right, like the true pixelized sky. Likewise, terms that involve i^n have a $\mathbf{J}^n$ matrix on the left.

In 2D, the situation is a bit more complicated. For clarity, let us treat $\mathbf{P}$ as a
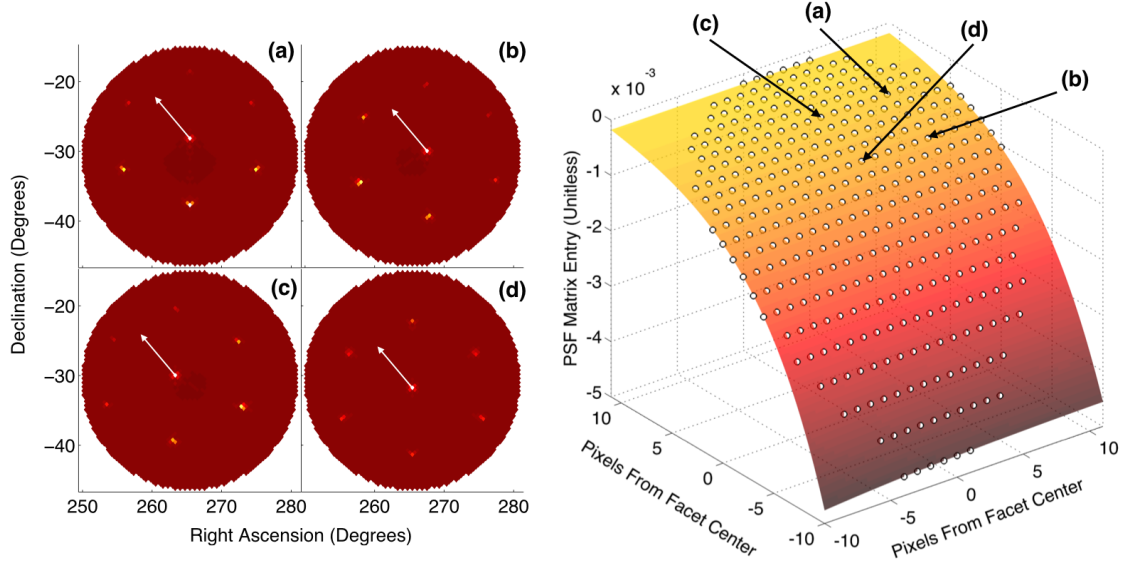
Figure 3-13: Though our point spread functions are not translationally invariant—a fact we saw clearly in Figure 3-3—their translational variation is fairly smooth and can be captured by a relatively low order polynomial. In this figure, we examine a typical example consisting of all the entries in **P** displaced exactly 15 pixels along one of the two principal axes of the pixelization from the center of the main lobe of the synthesized beam. This displacement is represented by the four identical white arrows on top of the point spread functions in the left-hand panel. All such entries in **P** (white circles in the right-hand planel) are plotted as a function of the displacement of the corresponding main lobe from the facet center. The points indicated by the white arrows in panels (a) through (d) are the same as the white circles indicated on the right hand plot. We then fit those points as a low-order 2D polynomial (in this case, as a quartic), which we plot as a colored plane cutting through them. The fit on the right hand side is merely one in a family of fits to each possible displacement vector from the main lobe of the PSF. Fitting the translational variation of the PSF in this way is potentially very useful, since a sparse representation of **P**, the matrix of point spread functions, would allow us to quickly multiply it by a vector. Though this is not important for mapmaking, it is important for estimating power spectra from the dirty maps and mapping statistics produced by our method.

4-indexed object, mapping two spatial dimensions to two other spatial dimensions. We approximate $\mathbf{P}$ as a polynomial sum of the form

$$P_{ii'jj'} = \sum_{n,m} t_{n,m,i-i',j-j'} (i+i')^n (j+j')^m. \tag{3.49}$$

Now $\mathbf{T}_{n,m}$ is a "block Toeplitz" matrix, essentially a Toeplitz matrix of Toeplitz matrices. Thankfully, multiplying by the matrix by a vector of size N_{PSF} still only scales as $\mathcal{O}(N_{\mathrm{PSF}} \log N_{\mathrm{PSF}})$ [115]. Expanding this to second order yields quite a few more terms:

$$\mathbf{P} \approx \overbrace{\mathbf{T}_{0,0}}^{0^{\mathrm{th}} \text{ Order}} +$$

$$\overbrace{\mathbf{T}_{1,0}(\mathbf{J} \otimes \mathbf{I}) + (\mathbf{J} \otimes \mathbf{I})\mathbf{T}_{1,0}}^{1^{\mathrm{st}} \text{ Order}} +$$

$$\mathbf{T}_{0,1}(\mathbf{I} \otimes \mathbf{J}) + (\mathbf{I} \otimes \mathbf{J})\mathbf{T}_{0,1} +$$

$$\overbrace{\mathbf{T}_{2,0}(\mathbf{J}^2 \otimes \mathbf{I}) + 2(\mathbf{J} \otimes \mathbf{I})\mathbf{T}_{2,0}(\mathbf{J} \otimes \mathbf{I}) + (\mathbf{J}^2 \otimes \mathbf{I})\mathbf{T}_{2,0}}^{2^{\mathrm{nd}} \text{ Order}} +$$

$$\mathbf{T}_{1,1}(\mathbf{J} \otimes \mathbf{J}) + (\mathbf{J} \otimes \mathbf{I})\mathbf{T}_{1,1}(\mathbf{I} \otimes \mathbf{J}) +$$

$$(\mathbf{I} \otimes \mathbf{J})\mathbf{T}_{1,1}(\mathbf{J} \otimes \mathbf{I}) + (\mathbf{J} \otimes \mathbf{J})\mathbf{T}_{1,1} +$$

$$\mathbf{T}_{0,2}(\mathbf{I} \otimes \mathbf{J}^2) + 2(\mathbf{I} \otimes \mathbf{J})\mathbf{T}_{0,2}(\mathbf{I} \otimes \mathbf{J}) + (\mathbf{J}^2 \otimes \mathbf{I})\mathbf{T}_{0,2}. \tag{3.50}$$

Here, we adopt the convention that all tensor products have the matrices in the i or i' dimension on the left-hand side of the $\otimes$ symbol and j or j' matrices on the right-hand side. In fact, it turns out that the exact number of polynomial terms is

$$N_{\mathrm{poly}} = \frac{1}{24} \left(24 + 50\omega + 35\omega^2 + 10\omega^3 + \omega^4 \right), \tag{3.51}$$

where $\omega \equiv \max(n+m)$ is the highest order polynomial considered.

The good news is that this fitting works pretty well at relatively low order, such as cubic or quartic. In Figure 3-14 we calculate the relative error between a dirty map computed by convolving the pixelized "true" sky with a very accurate $\mathbf{P}$ (one
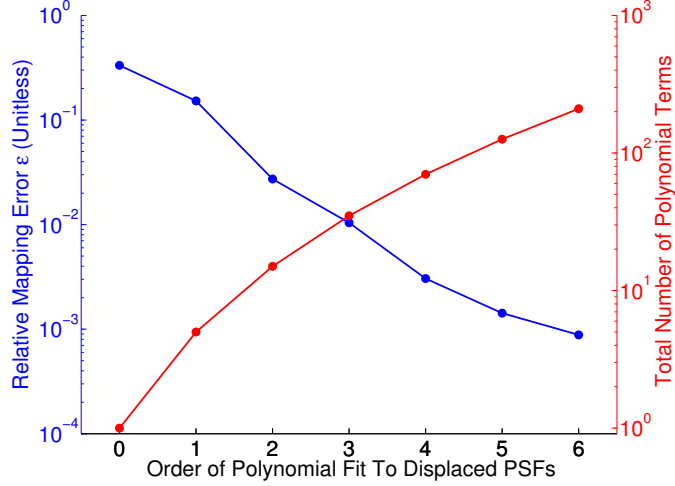
Figure 3-14: Approximating the translational variation of the point spread function with a low order polynomial can produce fairly small errors at a relatively low accuracy cost. Here we show the accuracy of multiplying a polynomially approximated **P** with the true sky compares to a direct calculation (using a large PSF truncation radius and no snapshotting). The errors are not negligible and the use of this approximation requires a carful examination of the accuracy requirements of the dirty maps. This technique saves time when the total number of terms in a polynomial/Toeplitz expansion of **P** is considerably smaller than the number of pixels in a facet. Unfortunately, that number of terms grows quartically with the polynomial order, meaning that very high orders and thus very high accuracy are not computationally useful.

computed with a large truncation radius and no snapshotting) and one computed with a polynomial fit to the translationally varying component of **P**. We find that the method outlined above can faithfully reproduce the dirty map to high precision.

Increasing accuracy, however, comes at a steep cost. While multiplication of **P** by a vector for a single frequency can be performed in $\mathcal{O}(N_{\mathrm{facet}} N_{\mathrm{PSF}})$, multiplication of a polynomially-approximated **P** takes $\mathcal{O}(N_{\mathrm{poly}} N_{\mathrm{PSF}} \log N_{\mathrm{PSF}})$. Since N_{poly} scales with the fourth power of the maximum order, it gets expensive very quickly. Thus the method outlined above is especially useful when $\sim 1\%$ to 0.1% errors are acceptable or when facets are exceptionally big or of exceptionally high resolution.

It is possible to reduce that cost by attacking the problem with a hybrid approach. We find that the biggest fitting errors come far from the facet center, especially in the brightest side lobes. This makes sense, since it is where the notion of a fixed "displacement" from the main lobe of the PSF runs up against the limits of the flat-

sky approximation. One could use this technique to incorporate the effects of most of $\mathbf{P}$, zeroing out the contributions from side lobe displacements. Then we could take the remainder of the $\mathbf{P}$ into account by simple matrix multiplication, achieving the same error with many fewer polynomial terms.

With big facets or at high resolution, PSF fitting serves another function. If the computational cost of mapmaking and power spectrum estimation is dominated by the matrix multiplication $\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{A}$ in the calculation of $\mathbf{P}$, we can choose to calculate only a representative sample of the entries in $\mathbf{P}$ (i.e. only some of the points on the right-hand side of Figure 3-13). Then we would rely on the fact that the polynomial fit is overdetermined to back out the missing entries.[12]

Whether or not to use the polynomial approximation to the $\mathbf{P}$ will depend on the exact telescope configuration and the nature of the mapmaking and power spectrum estimation problems at hand. If we want to try to precisely subtract foregrounds and work deep within the wedge, the polynomial approximation might not be good enough. However, if instead our power spectrum estimation strategy is to focus on isolating the EoR window and projecting out foreground-dominated modes entirely, it is less important that we very precisely understand the effect of the instrument. In that case, it is more likely that the polynomial PSF fitting approach outlined above will be useful. We explore these two approaches in the context of the mapmaking formalism in Appendix 3.B.

3.3.7   Computational Methods Summary

In the previous three sections, we explored three different ways of speeding up either the calculation of $\mathbf{P}$ or the multiplication of $\mathbf{P}$ by a vector. In Table 3.1 we summarize those results. In general, we find that PSF truncation and snapshotting have the most utility for HERA. PSF fitting, in the fiducial scenario we considered, is the least helpful. However, for a telescope with much higher angular resolution than HERA,

---

[12]It is worth noting that although a large number of terms might be needed to multiply $\mathbf{P}$ by a vector, there are not nearly so many free parameters in the fits. The number of free parameters needed to find a best-fit surface like that in Figure 3-13 only scales like the square of the highest polynomial order.

PSF fitting is likely to be more useful, since multiplication of a vector by $\mathbf{P}$ scales quadratically with the number of pixels in the facet.

While these results are specific to HERA, we can draw a few general conclusions. For HERA at 150 MHz, the first side lobes are about 13° from the main lobe of the synthesized beam. At 13° from the zenith, the primary beam is down by 20 dB. In general, it is likely we will only be able to truncate the PSF in regions where the primary beam is small, meaning that a telescope with a broader primary beam will benefit less from cropping in a way that scales quadratically with the PSF truncation radius and therefore also the PSF's full width at half maximum. By contrast, larger primary beams are more slowly varying spatially, meaning that longer snapshots are likely to achieve the same error. If the primary beam is relatively smooth, that benefit scales inverse-linearly with the size of the primary beam.

Though we used 1% as a somewhat arbitrary point of comparison in Table 3.1, it remains an open question how good our models of the $\mathbf{P}$ have to be. The only comprehensive way to answer this question is through a full end-to-end simulation of the signal, noise, and foregrounds all passed through a simulated instrument, a mapmaking code, and then power spectrum estimation. That sort of quantitative answer is outside the scope of this paper. However, it is worthwhile to enumerate the ways in which we need to use $\mathbf{P}$ to make maps and estimate power spectra and to examine the accuracy requirements for those tasks. By our count, $\mathbf{P}$ appears in six key places in the power spectrum estimation process:

1. When we calculate $\widehat{\mathbf{x}}$, we need $\mathbf{P}$ to define $\mathbf{D}$. However, looking closely at Equation (3.13) shows that $\mathbf{D}$ actually cancels out—the factor of $\mathbf{D}$ in each $\widehat{\mathbf{x}}$ and the two in $\mathbf{C}_{,\beta}$ are canceled by the two in each $\mathbf{C}^{-1}$. Therefore, it does not matter whether we get $\mathbf{D}$ right or not, as long as we are consistent about what we use for it. This makes sense, $\mathbf{D}$ was supposed to be an arbitrary choice, so as long as it is invertible, there is no way to get it "wrong" per se.

2. $\mathbf{P}$ also appears in our models for the parts of $\boldsymbol{\mu}$ and $\mathbf{C}^{FG}$ corresponding to bright point sources in $\mathbf{C}^{FG}$. Accounting properly for bright point sources has

Approximation Parameter	PSF Truncation Radius	Snapshot Time	PSF Fitting Order
Improves	Size of **P**	Steps in computing **P**	Multiplying by **P**
More exact when...	...larger	...smaller	...larger
Cost scaling	Quadratic	Inverse Linear	Quartic
Acheives 1% Error[a] for HERA at	5° or[b] 15°	10 minutes	3rd order
Speed-up at 1% error	~60 or ~500	~300	~5

Table 3.1: Summary of the techniques we use to approximate the calculation of multiplication by the matrix of point spread functions, **P**, in order to dramatically improve the speed of those operations. Faceting alone makes calculating **P** faster by a factor of 500 in our fiducial scenario. Combining PSF truncation and snapshotting brings the calculation of **P** well within the realm of feasibility. The benefit to fitting the PSF with polynomials and Toeplitz matrices is relatively small for our scenario, but it gets much better for higher resolution instruments.

---

[a] Assumes HEALPix $N_{\text{side}} = 256$, 2 s integrations, and 10° diameter facets.
[b] This depends on whether point sources are included, since they are mostly inside the facet in our simulations, depressing the error at small truncation radius.

the highest bang for the buck, in the sense that it is relatively straightforward to model both their means and covariances in the dirty map. In Section 3.3.2, we discussed how we could account for bright point sources with well-characterized positions, fluxes, and spectral indices by calculating a column in $\mathbf{P}$ that maps the point source to the entire facet in the dirty map. For that calculation, the PSF truncation radius is irrelevant because we account for the brightest sources in a separate part of the PSF independent of the HEALPix grid. Since we calculate only a moderate number of columns of $\mathbf{P}$, we do not even have to combine integrations into the snapshot. For bright point sources, it is not much extra effort to get $\mathbf{P}$ almost exactly right.

3. By contrast, diffuse emission from confusion-limited and galactic synchrotron emission in $\boldsymbol{\mu}$ and $\mathbf{C}^{FG}$ depends, as we have argued, on knowing how $\mathbf{P}$ maps a large part of the true sky onto the facet. It is in this context that approximate versions of $\mathbf{P}$ are the most useful, but also where they are potentially the most worrisome. Galactic and confusion-limited foregrounds are still orders of magnitude stronger than the cosmological signal and understanding them precisely is very important. Forming $\boldsymbol{\mu}$ from these foregrounds should be comparatively easy—all we need to do is take our sky model, compute visibilities, and then pass it through our mapmaking routine. We do not even need to calculate the full $\mathbf{P}$ matrix. Writing down $\mathbf{C}^{FG}$ is substantially more difficult, since $\mathbf{C}^{FG} = \mathbf{P}\mathbf{C}^{FG}_{\text{model}}\mathbf{P}^{\mathsf{T}}$. Exactly how well we need to know $\mathbf{P}$ in order for $\mathbf{C}^{FG}$ to accurately reflect the foreground uncertainty depends on the specific instrument, the foreground model, and our uncertainty about that model. A quantitative answer requires detailed covariance modeling outside the scope of this work and is therefore left for future investigation.

4. Modeling noise properly is extremely important since inside the EoR window only noise and signal should matter. A slight mismodeling of noise due to an error in the calculation of $\mathbf{C}^N$ could lead to an erroneous detection. If however we perform mapmaking twice from a cross power spectrum of interleaved

timesteps, we can eliminate noise bias [59, 173]. If we do that, it is acceptable (albeit not optimal) to be very conservative in our model of the instrumental noise, effectively increasing the error bars due to noise without biasing our measurement. If we adopt this conservative stance, then we can confidently use an approximate form of $\mathbf{P}$ when calculating $\mathbf{C}^N$.

5. Modeling $\mathbf{C}^S$ is mostly important for the calculation of sample variance. In any foreseeable experiment, this is a small contribution to the error. Getting $\mathbf{C}^S$ slightly wrong is unlikely to be the dominant error associated with approximating $\mathbf{P}$.

6. The $\mathbf{C}_{,\beta}$ family of matrices is necessary for telling us how to translate properly weighted dirty maps into power spectra. We need $\mathbf{P}$ to be as accurate as the precision with which we would like to measure the power spectrum.

In general, the question of exactly how accurately we need to know $\mathbf{P}$—and by extension, exactly how well we need to understand our instruments—is an open question for future investigation.

3.4 Summary and Future Directions

In this work, we showed how to make precise maps with well-understood statistics specifically for $21\,\mathrm{cm}$ power spectrum estimation. We investigated how to connect the framework of optimal mapmaking to that of inverse-variance weighted quadratic power spectrum estimation in order to understand what sort of maps and map statistics we need for power spectrum estimation. We showed that in addition to the dirty map estimator $\widehat{\mathbf{x}}$, we need the matrix of point spread functions, $\mathbf{P}$, and the noise covariance matrix which takes a gratifyingly simple form: $\mathbf{C}^N = \mathbf{P}\mathbf{D}^\mathsf{T}$ where $\mathbf{D}$ is an invertible normalization matrix that we can choose to be diagonal.

This analysis technology will allow us to consistently integrate our best understanding of an instrument with our best models for noise, foregrounds, and the cosmological signal. Not only does this approach help prevent the loss of cosmological

information, but it will allow for a precise measurement of the 21 cm power spectrum and for the confident and robust description of the errors in our estimates.

In the main part of this work, we focused on the matrix of point spread functions, $\mathbf{P}$, which relates the true sky to our dirty maps. We calculated simulated dirty maps and PSFs for HERA, the upcoming Hydrogen Epoch of Reionization Array. While calculating $\mathbf{P}$ exactly is computationally prohibitive, we explored three methods for approximating $\mathbf{P}$. First, we explored how making maps in facets with truncated PSFs can dramatically reduce the computational cost of calculation $\mathbf{P}$ for only a small hit to accuracy. Next we showed how to combine consecutive integrations while controlling for the errors introduced by the process. It turns out that observations many minutes apart can be combined with minimal error. Lastly, we showed how the multiplication of $\mathbf{P}$ by a vector—a necessary step for power spectrum estimation—might be sped up by approximating its translational variance as slowly varying. Though the cost scaling of this approximation is steep, we find this technique especially promising when moderate errors are tolerable or for instruments with high angular resolution.

Just as importantly, all these methods have tunable knobs—they can be made more accurate at the cost of speed or memory. Though our specific, quantitative results are only applicable to HERA, the accuracy trade-offs and the computational scalings we find should be quite general. In that sense, we hope that this work serves as a versatile guide to mapmaking in the context of 21 cm cosmology.

Much work remains to be done to develop a clear and computationally tractable pathway from visibilities all the way to power spectra with rigorous errors and error correlations. Even after connecting this work to an appropriately updated version of the Dillon et al. [58] algorithm, one still needs to assess the effect of our approximations, as well as a number of important data analysis choices, on power spectrum estimates and ultimately on cosmological parameter constraints. Though the errors incurred by each can be made arbitrarily small, it is difficult to say yet what level of approximation is tolerable. This is an open question for future work.

We would like to see a full end-to-end simulation, starting with the 21 cm signal, passing through the instrument, and ending with power spectra and their statistics.

Such a full-scale test could prove the effectiveness of these techniques and clarify exactly what the approximations utilized both in this work and in Dillon et al. [58] do to our measured power spectra. A power spectrum estimation technique that passes such a test with realistic foregrounds and noise will be the one to produce trustworthy cosmological measurements.

3.A   Appendix: Polarization and Heterogenous Primary Beams

In Section 3.2.1, we worked out the relationship between visibilities and the true sky in terms of the matrix $\mathbf{A}$. For the sake of simplicity, we made two assumptions that, in this appendix, we would like to relax.

First, we ignored the effect of polarization. Though the 21 cm signal is unpolarized, astrophysical foregrounds are generally polarized. And because the primary beams of the two orthogonal polarizations measured by a single element are different, the polarization of sources is important. This is especially important for sources with high rotation measures [150]. Second, we ignored the possibility that not every element has the same primary beam. It is possible that an array is intentionally constructed with multiple kinds of elements. It is also generally true that different elements will behave slightly differently, just due to the variations in their construction. If we are able to measure that variation—which is no small task—we would like to take it into account.

Let us begin with polarization. There are a number of different conventions for expressing polarization [210], but one relatively straightforward one is to replace $I(\hat{\mathbf{r}}, \nu)$ with a four-element vector $\mathbf{I}(\hat{\mathbf{r}}, \nu)$ containing Stokes I, Q, U, and V parameters. Instead of one visibility per baseline and frequency, we now measure four, one for each of the pairs of polarizations of antennas, xx, xy, yx, and yy. In this case, $B(\hat{\mathbf{r}}, \nu)$ becomes $\mathbf{B}(\hat{\mathbf{r}}, \nu)$, a 4×4 matrix that describes the response of each type of visibility to each polarization and direction.

Otherwise, not much changes. The sky vector we are estimating gets four times bigger and the number of visibilities also gets four times bigger (though there are simplifications in practice, since xy and yx visibilities are just complex conjugates of one another and can be averaged together to reduce noise). The $\mathbf{A}$ matrix is not fundamentally different. Though it may seem like this makes the problem of computing $\mathbf{P}$ 64 times harder, that is fortunately not the case.

Fundamentally, we want to estimate the cosmological signal from our best guess at the Stokes I map. Foregrounds can have I, Q, and U components—astrophysical sources are not circularly polarized. So what we really want is a $\mathbf{P}$ matrix that maps I, Q, and U on the true sky, through xx and yy visibilities, to a dirty map of Stokes I. That is only six times more difficult than the calculations outlined above. If we do not want to model our foreground residual as polarized, then $\mathbf{P}$ is only twice as complicated as before—we just need to calculate $\boldsymbol{\mu}$ through a more complicated mapmaking procedure involving the expanded definition of $\mathbf{A}$.

The issue with polarization is in many ways similar to the problem of heterogeneous primary beams. After all, the two polarization's dipoles generally have two different primary beams. Since the calculation of $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$ is the computationally limiting step in our method, it is not significantly more difficult to treat multiple kinds of primary beam products $B(\hat{\mathbf{r}}, \nu)$ when calculating $\mathbf{A}$, each row having a potentially different $B(\hat{\mathbf{r}}, \nu)$. This gives us a straightforward way to account for arrays that include multiple types of antenna elements.

Of greater concern is the fact that every element in a real array has a slightly different beam—even if it was designed to be homogenous. For a minimally redundant array, this does not matter. If we know the correct primary beam for every antenna, we can write down $\mathbf{A}$ exactly. For a highly redundant array like HERA, antenna heterogeneity breaks the redundancy of baselines. If we want to include all measured visibilities in our maps, we may need to treat visibilities involving the most discrepant antennas separately. If we had to go further and treat every visibility separately, that would make $\mathbf{P}$ two orders of magnitude more difficult to calculate for HERA. If we can measure primary beams for all of our antennas, it would be worthwhile to simulate

the error associated with the approximation that they are all the same. This is left to future work. Fortunately, it is theoretically possible to take into account slight variations between elements in the framework we have outlined.

3.B Appendix: A Foreground Avoidance Approach to Power Spectrum Estimation

The power spectrum estimation method we outlined in Section 3.2.3 is a promising way to enlarge the EoR window and gain the additional sensitivity forecasted by [184]. However, it is not the simplest approach. Instead of directly modeling foregrounds, we could choose to simply throw out all the modes that we believe to be foreground contaminated. The foreground avoidance approach was pioneered by [172] and used to produce the best current limits on the 21 cm power spectrum by [173] and [105]. This choice should be more robust to foreground mismodeling than subtraction, since we are merely trying to isolate foreground free regions of Fourier space from the effects of regions we have given up on. Where exactly we draw the line between wedge and window is a question that deserves further investigation with both simulations and real data.

One might ask why foreground avoidance estimators are interesting when the whole point of making maps like ours was to compress the data in a space where foregrounds were most naturally subtracted. There are a few reasons. First, foreground avoidance is simpler than foreground subtraction. If we are going to try to subtract foregrounds, it is worthwhile to first perform the simpler, more robust procedure so we have a baseline for comparison. Second, even if we are only interested in mitigating the effect of foregrounds by avoiding them, this method gives a proper accounting for $\mathbf{C}^N$, $\mathbf{C}^S$, and $\mathbf{C}_{,\alpha}$, without making any of the approximations previously relied upon about there being no correlations between uv cells or that uniform weighted maps have no PSF. Third, the technique is fairly directly comparable to that of [172] without the additional assumption that delay modes for a given visibility map neatly

to band powers or the computational challenges of [125, 126]. And finally, we may also want to implement a hybrid approach, similar in spirit to [173], where we project out modes deep into the wedge but try to subtract foregrounds nearer the edge of the wedge.[13]

Therefore, it is worthwhile to write down the general framework for foreground avoidance in the context of optimal mapmaking. The idea is relatively simple. Let's define a new dirty map estimator, $\widehat{\mathbf{x}}'$, defined as

$$\widehat{\mathbf{x}}' \equiv \mathbf{\Pi}\widehat{\mathbf{x}} \tag{3.52}$$

where $\mathbf{\Pi}$ is a projection matrix that has eigenvalues of 0 or 1 only. As with all projection matrices, $\mathbf{\Pi} = \mathbf{\Pi}^{\mathsf{T}} = \mathbf{\Pi}^2$. The matrix $\mathbf{\Pi}$ Fourier transforms the data cube, sets all modes outside the EoR window to zero, and Fourier transforms back. It also means that we need to replace $\mathbf{C}$ with $\mathbf{C}'$ where

$$\mathbf{C}' = \mathbf{\Pi}\mathbf{C}\mathbf{\Pi}. \tag{3.53}$$

By construction, the projection eliminates the foregrounds in μ, meaning that

$$\mathbf{\Pi}\langle\widehat{\mathbf{x}}\rangle = \mathbf{\Pi}\mu \approx 0. \tag{3.54}$$

Likewise, the part of the covariance associated with the foregrounds should also go to zero. Hence,

$$\mathbf{\Pi}\mathbf{C}^{FG}\mathbf{\Pi} \approx 0, \tag{3.55}$$

which means that

$$\mathbf{C}' = \mathbf{\Pi}\left[\mathbf{C}^S + \mathbf{C}^N\right]\mathbf{\Pi}. \tag{3.56}$$

---

[13]This is similar in spirit to what WMAP did [91]. They first masked out the galaxy and the brightest point sources, then they performed foreground subtraction in the map and foreground residual bias subtraction in the angular power spectrum. For us, the major difference is that the we do both

This also changes $\mathbf{C}_{,\alpha}$ which now takes the form

$$\mathbf{C}'_{,\alpha} = \mathbf{\Pi}\mathbf{C}_{,\alpha}\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}\mathbf{Q}_{\alpha}\mathbf{P}^{\mathsf{T}}\mathbf{\Pi}. \tag{3.57}$$

Of course, the new covariance has many zero eigenvalues, which means that it is not invertible. That is not a problem since we can replace $(\mathbf{C}')^{-1}$ by its "pseudoinverse" [213], defined as

$$(\mathbf{C}')^{-1}_{\text{psuedo}} = \mathbf{\Pi}\left[\mathbf{\Pi}\mathbf{C}'\mathbf{\Pi} + \gamma(\mathbf{I} - \mathbf{\Pi})\right]^{-1}\mathbf{\Pi} \tag{3.58}$$

where γ can be any (numerically reasonable) nonzero number without changing the result. The pseudoinverse reflects the idea that we want to completely throw out any power in possibly foreground-contaminated modes but also that we want to express infinite uncertainty in the modes—in other words, to give them no weight. This will accurately account for the fact that we have no information about these modes.

Putting all that together, our new quadratic estimator $\widehat{\mathbf{p}}$ is

$$\widehat{p}_{\alpha} = \frac{1}{2}M_{\alpha\beta}\widehat{\mathbf{x}}^{\mathsf{T}}(\mathbf{C}')^{-1}_{\text{psuedo}}\mathbf{P}\mathbf{Q}_{\beta}\mathbf{P}^{\mathsf{T}}(\mathbf{C}')^{-1}_{\text{psuedo}}\widehat{\mathbf{x}} - b_{\alpha}, \tag{3.59}$$

where we have used the fact that $\mathbf{\Pi}^{2} = \mathbf{\Pi}$. The estimator is not lossless, but it can still be unbiased in the region of Fourier space not projected out and have rigorously defined and calculable error properties.

Part II

Early Results from New Telescopes

Chapter 4

Overcoming Real-World Obstacles in 21 cm Power Spectrum Estimation: A Method Demonstration and Results from Early Murchison Widefield Array Data

The content of this chapter was submitted to Physical Review D *on April 25, 2013 and published [59] as* Overcoming real-world obstacles in 21 cm power spectrum estimation: A method demonstration and results from early Murchison Widefield Array data *on January 15, 2014.*

4.1 Introduction

In recent years, 21 cm tomography has emerged as a promising probe of the Epoch of Reionization (EoR). As a direct measurement of the three-dimensional distribution of neutral hydrogen at high redshift, the technique will allow detailed study of the complex astrophysical interplay between the intergalactic medium and the first luminous structures of our Universe. This will eventually pave the way towards the use of

21 cm tomography to constrain cosmological parameters to exquisite precision, thanks to the enormity of the physical space within its reach (please see, e.g., Furlanetto et al. [74], Morales and Wyithe [154], Pritchard and Loeb [188], Loeb and Furlanetto [127] for recent reviews).

To date, observational efforts have focused on measurements of the 21 cm power spectrum. Such a measurement is exceedingly difficult. Sensitivity requirements are extreme, requiring thousands of hours of integration and large collecting areas [151, 25, 117, 87, 170]. Adding to this challenge is the fact that raw sensitivity is insufficient—what counts is sensitivity to the cosmological signal above expected contaminants like galactic synchrotron radiation, which are three to four orders of magnitude brighter at the relevant frequencies [53, 107, 18, 182].

To deal with these challenges, numerous techniques have been proposed and implemented for foreground mitigation and power spectrum estimation. These include foreground removal via parametric fits [231, 28, 122, 123], non-parametric methods [86, 39, 40], principal component analyses [166, 121, 137, 167], filtering [78, 176, 172], frequency stacking [41], and quadratic methods [120, 58, 197]. In almost all of these proposals, foregrounds are separated from the cosmological signal by taking advantage of the differences in their spectra. Foregrounds are dominated by continuum processes and thus have smooth spectra. On the other hand, because the cosmological line-of-sight distance maps to the observed frequency of the redshifted 21 cm line, the rapid fluctuations in the brightness temperature distribution that are expected from theory will map to a measured cosmological signal with jagged, rapidly fluctuating spectra. When these spectral differences are considered in conjunction with instrumental characteristics, one can identify an "EoR window": a region in Fourier space where power spectrum measurements are expected to be relatively free from foregrounds [50, 172, 230, 156, 225, 218]. This is shown schematically in Figure 4-1, where we have used early Murchison Widefield Array (MWA) data to estimate the power spectrum as a function of $k_\perp$ (Fourier mode perpendicular to the line-of-sight) and $k_\parallel$ (Fourier mode parallel to the line-of-sight). More details regarding this figuew are provided in Section 4.3; for now we simply wish to draw attention to the existence
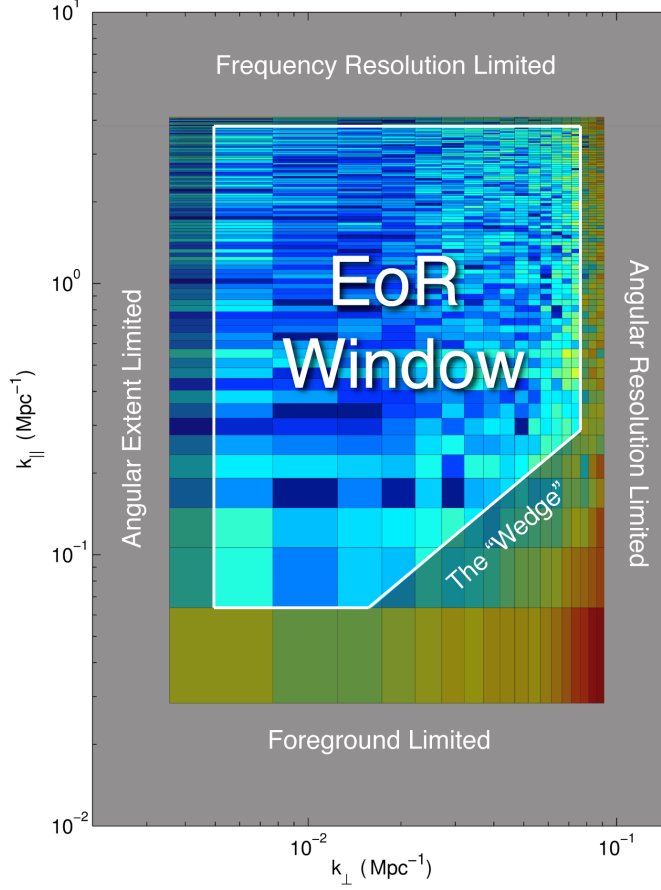
Figure 4-1: .
The "EoR window," a region of Fourier space with relatively low noise and foregrounds, is thought to present the best opportunity for measuring the cosmological 21 cm power spectrum during the Epoch of Reionization. Here we show an example power spectrum from early MWA data, as a function of $k_\perp$ (Fourier components perpendicular to the line of sight) and $k_\parallel$ (Fourier components parallel to the line of sight). More details on how we have calculated and plotted $P(k_\perp, k_\parallel)$ are found in Section 4.3. We schematically highlight the instrumental and foreground effects that that delimit the EoR window—the coldest part of this power spectrum. At low and high $k_\perp$, measurements are limited by an instrument's ability to probe the largest and smallest angular scales, respectively. Limited spectral resolution causes similar effects at the highest $k_\parallel$. As spectrally smooth sources, foregrounds inhabit primarily the low $k_\parallel$ regions. Thanks to chromatic instrumental effects, however, there is a slight encroachment of foregrounds towards higher $k_\parallel$ at higher $k_\perp$, in what has been colloquially termed the "wedge" feature.

of a relatively contaminant-free region in the middle of the $k_\perp$-$k_\parallel$ plane. This clean region is what we denote the EoR window.

The EoR window is generally considered the sweet spot for an initial detection of

205

the cosmological 21 cm power spectrum, and constraints are likely to degrade away from the window. At high $k_\perp$ (i.e., the finest angular features on the sky), errors increase due to the angular resolution limitations of one's instrument. For an interferometer, this resolution is roughly set by the length of the longest baseline. Conversely, the shortest baselines define the largest modes that are observable by the instrument. Errors therefore also increase at the lowest $k_\perp$ where again there are few baselines.

A similar limitation defines the boundary of the EoR window at high $k_\parallel$. Since the spectral nature of 21 cm measurements mean that different observed frequencies map to different redshifts, the highest $k_\parallel$ modes are inaccessible due to the limited spectral resolution of one's instrument. At low $k_\parallel$, one probes spectrally smooth modes—precisely those that are expected to be foreground contaminated. Thus there is another boundary to the EoR window at low $k_\parallel$.

A final delineation of the EoR window is provided by the region labeled as the "wedge" in Figure 4-1. The wedge feature is a result of an interplay between angular and spectral effects. Simulations have shown that the wedge is the effect of chromaticity in one's synthesized beam (which is inevitable when an interferometer is used to survey the sky). This chromaticity imprints unsmooth spectral features on measured foregrounds, resulting in foreground contamination beyond the lowest $k_\parallel$ modes even if the foregrounds themselves are spectrally smooth. Luckily, this sort of additional contamination follows a reasonably predictable pattern in the $k_\perp$-$k_\parallel$ plane, and in the limit of intrinsically smooth foregrounds, the wedge can be shown to extend no farther than the line

$$k_\parallel = \left[\sin\theta_{\text{field}} \frac{D_M(z)E(z)}{D_H(1+z)} \right] k_\perp, \tag{4.1}$$

where $D_H \equiv c/H_0$, $E(z) \equiv \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$, $D_M(z) \equiv \int_0^z dz'/E(z')$, θ_{field} is angular radius of the the field-of-view, and c, H_0, Ω_m, and Ω_Λ have their usual meanings [50, 230, 156, 225]. Intuitively, the foreground-contaminated wedge extends to higher $k_\parallel$ at higher $k_\perp$ because the high $k_\perp$ modes are probed by the longer baselines of an interferometer array, which have higher fringe rates that more effectively imprint

spectral structure in the measured signals. For an alternate but equivalent explanation in terms of delay modes, please see the illuminating discussion in Parsons et al. [172].

The concept of an EoR window is important in that it provides relatively strict boundaries that separate fairly foreground-free regions of Fourier space from heavily foreground-contaminated ones. It therefore provides one with the option of practicing foreground avoidance rather than foreground subtraction. If it turns out that foregrounds cannot be modeled well enough to be directly subtracted with the level of precision required to detect the cosmological signal, foreground avoidance becomes an important alternative, in that the only way to robustly suppress foregrounds is to preferentially make measurements within the EoR window. Likely, some combination of the two strategies—foreground subtraction and foreground avoidance—will prove useful for the detection of the 21 cm power spectrum. Of course, measurements within the EoR window are still contaminated by instrumental noise, but fortunately the noise integrates down with further observation time (as long as calibration errors and other instrumental systematics can be sufficiently minimized). Observationally, it is encouraging that the EoR window has now been shown to be free of foregrounds to better than one part in a hundred in power [182].

As experimental sensitivities increase, however, one must take care to preserve the cleanliness of the EoR window to an even higher dynamic range. There are several ways in which our notion of the EoR window may be compromised. First, as experiments integrate in time and acquire greater sensitivity, we may discover that our approximation of spectrally smooth foregrounds is insufficiently good for a detection of the (faint) cosmological signal. In other words, foreground sources may have small but non-negligible high $k_{\parallel}$ components in their spectra that have thus far gone undetected. This would translate into a smaller-than-expected EoR window. In addition, even intrinsically smooth foregrounds may appear jagged in a real measurement because of instrumental effects such as imperfect calibration. The precise interferometer layout may also result in unsmooth artifacts that arise from combining data from non-redundant baselines [89]. Finally, suppose that the aforementioned

effects are negligible and that the assumption of spectrally smooth foreground emission continues to hold. The EoR window still cannot be taken for granted because non-optimal data analysis techniques may result in unwanted foreground artifacts in the region. For the EoR window to exist at all, it is essential that power spectra are estimated in a rigorous fashion, with well-understood statistics.

The goal of this paper is to minimize unwanted data analysis artifacts by establishing methods for power spectrum estimation that are both robust and as optimal as possible. Previous efforts have rarely met both criteria: either the methods are robustly applicable to data with real-world artifacts but fail to achieve optimized (or even rigorously computable) error properties, or provide an optimal framework but ignore real-world complications. In this paper we extend the rigorous framework described in Liu and Tegmark [120] and Dillon et al. [58] to deal with real-world effects. The result is a computationally feasible approach to analyzing real data that not only preserves the cleanliness of the EoR window, but also rigorously keeps track of all relevant error statistics.

To demonstrate the applicability of our approach, we apply our techniques to early data from the Murchison Widefield Array (MWA). These data were derived from ~ 22 hours of tracked observations using an early, 32-element prototype array. The results are therefore not designed to be cosmologically competitive, but instead illustrate the rigor that will be required for an eventual detection of the EoR while also providing new measurements on the "wedge" feature that delineates the EoR window.

This paper is organized as follows. In Section 4.2 we discuss various real-world obstacles that must be dealt with when analyzing real data, and how one can overcome them while maintaining statistical rigor. We then apply our methods to MWA data in Section 4.3 as a "worked example", highlighting the importance of various subtleties of power spectrum estimation. In Section 4.4 we present some results from the data, emphasizing the agreement between theoretical expectations and our observations of the foreground wedge (particularly regarding the frequency dependence of the wedge). We also present upper limits on the cosmological 21 cm power spectrum over the broad

redshift range of $z = 6.2$ to $z = 11.1$. Finally, we summarize our conclusions in Section 4.5.

4.2 Systematic Methods for Dealing with Real-World Obstacles

To understand the gap between an analysis framework for idealized observations and any real-world data set, we enumerate and address six different obstacles that rather universally affect real data. Our goal in this section is to meet the challenges presented by these obstacles while maintaining as many of the advantages of the optimal framework as possible, which we reiterate in Section 4.2.1, especially the ability to minimize and precisely quantify the uncertainties in the measurements. In the following sections, we address the problems presented by large data volumes (Section 4.2.2), uncertainties in the properties of contaminants such as foregrounds (Section 4.2.3), incomplete uv coverage (Section 4.2.4), radio frequency interference (RFI) flagging (Section 4.2.5), foreground leakage into the EoR window (Section 4.2.6), and binning to spherically averaged power spectra (Section 4.2.7).

4.2.1 A Systematic Framework for Analyzing Idealized Observations

In this section, we briefly review the formalism of Liu and Tegmark [120] for optimal power spectrum estimation, which was adapted for 21 cm tomography from similar techniques used in galaxy survey and cosmic microwave background analysis [208, 23, 213, 215]. For now, we do not include real-world effects such as missing data from RFI flagging, and the purpose of later sections is to extend the formalism to take into account these complications.

In 21 cm tomography, one typically wishes to measure both the spherically-binned

power spectrum $P_{\text{sph}}(k)$, defined by

$$\langle \widetilde{T}^*(\mathbf{k})\widetilde{T}(\mathbf{k}')\rangle \equiv (2\pi)^3 P_{\text{sph}}(k)\delta(\mathbf{k}-\mathbf{k}'), \tag{4.2}$$

and the cylindrically-binned power spectrum $P_{\text{cyl}}(k_\perp, k_\parallel)$, defined by

$$\langle \widetilde{T}^*(\mathbf{k})\widetilde{T}(\mathbf{k}')\rangle \equiv (2\pi)^3 P_{\text{cyl}}(k_\perp, k_\parallel)\delta(\mathbf{k}-\mathbf{k}'), \tag{4.3}$$

with $\widetilde{T}(\mathbf{k})$ signifying the spatial Fourier transform of the 21 cm brightness temperature field $T(\mathbf{r})$, $\mathbf{k}$ denoting the spatial wavevector with magnitude k, and components $k_\perp$ and $k_\parallel$ as the components perpendicular and parallel to the line-of-sight, respectively. The angled brackets $\langle \cdots \rangle$ represent an ensemble average. The spherical power spectrum is useful for comparing to theoretical models, since it is obtained by angularly averaging over spherical shells in Fourier space, and thus makes the cosmologically relevant assumption of isotropy. The cylindrical power spectrum is useful for identifying instrumental and foreground effects, which possess a cylindrical symmetry rather than a spherical one. Typically, the cylindrical power spectrum is produced first as a tool for foreground isolation (i.e., to identify the EoR window), and then subsequently binned into a spherical power spectrum. This section concerns the estimation of the cylindrical power spectrum. Optimal binning techniques to go from the cylindrical spectrum to the spherical spectrum are discussed in Section 4.2.7.

In estimating a power spectrum from data, one must necessarily discretize the problem. We make the approximation that the power spectra are piecewise constant functions, such that we can describe them in terms of a vector of bandpowers with components p^α, where

$$p_\alpha \equiv P_{\text{cyl}}(k_\perp^\alpha, k_\parallel^\alpha). \tag{4.4}$$

It is the bandpowers and their error properties that one wishes to estimate from the data, which come in the form of a data vector $\mathbf{x}$. Intuitively, one can think of the data vector as a list of the 21 cm brightness temperatures measured at various locations in a three-dimensional "data cube". Rigorously, we define each element of the data

vector (i.e., each voxel of the data cube) as

$$\mathbf{x}_i \equiv \int T(\mathbf{r})\psi_i(\mathbf{r})d^3\mathbf{r}, \tag{4.5}$$

with $\psi_i(\mathbf{r})$ being the pixelization kernel and $T(\mathbf{r})$ as the (continuous) three-dimensional 21 cm brightness temperature field[1]. In this paper we take the ith pixelization kernel $\psi_i(\mathbf{r})$ to be a boxcar function centered on the ith voxel of the data.[2]

To estimate the αth bandpower from the data vector, we first form a quadratic estimator of the form

$$q_\alpha \equiv \frac{1}{2}(\mathbf{x} - \mathbf{m})^t \mathbf{C}^{-1} \mathbf{C}_{,\alpha} \mathbf{C}^{-1}(\mathbf{x} - \mathbf{m}) - \frac{1}{2}\mathrm{tr}[\mathbf{C}_{\mathrm{junk}}\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}], \tag{4.6}$$

where $\mathbf{m} \equiv \langle\mathbf{x}\rangle$ is the mean of the data, $\mathbf{C} \equiv \langle\mathbf{x}\mathbf{x}^t\rangle - \langle\mathbf{x}\rangle\langle\mathbf{x}\rangle^t$ is its covariance, $\mathbf{C}_{\mathrm{junk}}$ is the component of the covariance "junk"/contaminants (to be defined in the following section), and $\mathbf{C}_{,\alpha}$ is the derivative of the covariance with respect to the αth bandpower. Since we are approximating the power spectrum as piecewise constant, we have

$$\mathbf{C} = \mathbf{C}_{\mathrm{junk}} + \sum_\alpha p_\alpha \mathbf{C}_{,\alpha}. \tag{4.7}$$

Combined with Equation (4.5), this expression can be used to derive explicit forms for $\mathbf{C}_{,\alpha}$, which reveals that the matrix essentially Fourier transforms and bins the data [120, 58]. Intuitively, $\mathbf{C}_{,\alpha}$ can be thought of as the response in the data covariance $\mathbf{C}$ to the bandpower p_α. Thus, as long as one selects an appropriate form for $\mathbf{C}_{,\alpha}$, the formalism of this section can also be used to directly measure the spherical power spectrum. However, as we discussed above, in this paper we choose to first estimate the cylindrical power spectrum as an intermediate diagnostic step, to quantify and

---

[1] Of course, instrumental noise and foregrounds do not properly reside in a cosmological three-dimensional volume: noise is introduced in the electronics of the system, whereas foregrounds are "nearby" and only appear in the same location in the data cube as our cosmological signal by virtue of their frequency dependence. However, there is a gain in convenience and no loss of generality in assigning a noise and foreground contribution to each voxel, pretending that those contaminants also live in the observed cosmological volume.

[2] This choice, following [58], is motivated by the fact that the covariance between each pixel in this basis for both noise and foregrounds can be written in an algorithmically convenient way.

mitigate foregrounds better.

Once the q_αs have been formed, they need to be normalized using a suitable invertible matrix $\mathbf{M}$ to form the final bandpower estimates:

$$\widehat{\mathbf{p}} = \mathbf{M}\mathbf{q}, \tag{4.8}$$

where we have grouped the bandpower estimates $\widehat{p}_\alpha$ into a vector $\widehat{\mathbf{p}}$ (and similarly grouped the coefficients q_α and $\mathbf{q}$), with the hat (^) signifying the fact that we have formed an *estimator* of the true bandpowers[3]. We shall discuss different choices of $\mathbf{M}$ in Section 4.2.6.

To understand the uncertainty in our estimates, we compute several error properties. The first is the covariance matrix of the final measured bandpowers:

$$\boldsymbol{\Sigma} \equiv \langle \widehat{\mathbf{p}}\widehat{\mathbf{p}}^t \rangle - \langle \widehat{\mathbf{p}} \rangle \langle \widehat{\mathbf{p}} \rangle^t = \mathbf{M}\mathbf{F}\mathbf{M}^t, \tag{4.9}$$

where we have introduced the Fisher matrix $\mathbf{F}$, which has components

$$F_{\alpha\beta} = \frac{1}{2}\mathrm{tr}[\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}]. \tag{4.10}$$

The Fisher matrix also allows us to relate our estimated bandpowers $\widehat{\mathbf{p}}$ to the true bandpowers $\mathbf{p}$ via the window function matrix $\mathbf{W}$:

$$\langle \widehat{\mathbf{p}} \rangle = \mathbf{W}\mathbf{p}, \tag{4.11}$$

where $\mathbf{W}$ can be shown to take the form

$$\mathbf{W} = \mathbf{M}\mathbf{F}. \tag{4.12}$$

If we choose $\mathbf{M}$ such that the rows of $\mathbf{W}$ each sum to unity, Equation (4.11) shows that each bandpower estimate can be thought of as a weighted average of the truth,

---

[3]Note that $\mathbf{q}$, $\widehat{\mathbf{p}}$, and $\mathbf{M}$ live in a different vector space than $\mathbf{x}$, $\mathbf{C}$, and $\mathbf{C}_{,\alpha}$. The former are in a vector space where each component refers to a different bandpower, whereas the latter are in one where different components refer to different voxels.

with weights given by each row (each window function). Even with this normalization requirement, there are still many choices for $\mathbf{M}$. We discuss the various options and tradeoffs in Section 4.2.6.

Whatever the choice of $\mathbf{M}$, our estimator has optimal error properties in the sense that if $\widehat{\mathbf{p}}$ in Equation (4.11) is used to constrain parameters in some theoretical model, those measured parameters will have the smallest possible error bars given the observed data [208]. Our goal in the following sections will be to ensure that both these small error bars and our ability to rigorously compute them are preserved in the face of real-world difficulties.

4.2.2   A Real-World Obstacle: Data Volume

Perhaps the most glaring difficulty presented by the ideal technique outlined above is its computational cost. Much of that cost arises from the inversion of the data covariance matrix $\mathbf{C}$ in Equations (4.6) and (4.10), in addition to the multiplication of $\mathbf{C}$ and matrices of the same size. Both of these operations scale like $\mathcal{O}(N^3)$, where N is the number of voxels in each data vector. The computational cost makes taking full advantage of current generational interferometric data prohibitive, not to mention upcoming observational efforts that expect to produce 10^6 or more voxels of data.

One would like to retain the information theoretic advantages of the quadratic estimator method and its ability to precisely model errors and window functions, without $\mathcal{O}(N^3)$ complexity. The solution to this problem, developed and demonstrated in [58], comes from taking advantage of a number of symmetries and approximate symmetries of the survey geometry and the covariance matrix, $\mathbf{C}$, and can accelerate the technique to $\mathcal{O}(N \log N)$.

The fast method relies on assembling the data into a data cube with rectilinear voxels amenable to manipulation with the Fast Fourier Transform. This is equivalent to the assertion that each voxel represents an equal volume of comoving space, an approximation that relies on two restrictions on the data cube geometry. First, the range of frequencies considered must be small enough that $D_c(z)$ (the line-of-sight comoving distance, equal to $D_M(z)$ above in a spatially flat universe) is linear with

ν. Generally, one should limit oneself to analyzing the power spectrum of redshift ranges short enough that the evolution of the power spectrum during reionization can be neglected. This range, suggested by [135] to be $\Delta z \lesssim 0.5$, makes the approximation of a linear relationship between ν and $D_c(z)$ better than one part in 10^3 at the redshifts of interest to 21 cm cosmology.

Second, the assumption of equal volume voxels relies on the flat sky approximation. To achieve this the area surveyed can be broken into a number of subfields, each a few degrees on a side, for which the curvature of the sky can be neglected. As long as the angular extent of the data cube is smaller than $\sim 10°$, the flat sky approximation is correct to a few parts in 10^3.

By analyzing a rectilinear volume of the universe, all steps in calculating the band powers q_α can be performed quickly by exploiting various symmetries and taking advantage of the Fast Fourier Transform. The model for $\mathbf{C}$ can be broken up into a number of independent matrices representing signal, noise, and foregrounds. Each of these models, developed by [120], is well approximated by a sparse matrix in a convenient combination of real and Fourier spaces [58]. As a result, multiplication of a vector by $\mathbf{C}$ can be performed in $\mathcal{O}(N \log N)$. Dillon et al. [58] showed how that speed-up can be parlayed into a method for quickly calculating q_α using the Conjugate Gradient Method. The rapid convergence of the iterative method for calculating $\mathbf{C}^{-1}\mathbf{x}$ can be ensured by the application of a preconditioner which relies on the spectral smoothness of foregrounds and the fact that they are well described by only a few eigenmodes [121]. Then, by randomly simulating many data vectors from the covariance $\mathbf{C}$ and calculating q_α from each, the Fisher matrix can be estimated from the fact that

$$\mathbf{F} = \langle \mathbf{q}\mathbf{q}^t \rangle - \langle \mathbf{q} \rangle \langle \mathbf{q}^t \rangle, \tag{4.13}$$

which follows from Equation (4.9). All of this together allows for fast, optimal power spectrum estimation—including error bars and window functions—despite the challenge presented by an enormous volume of data.

4.2.3 A Real-World Obstacle: Uncertain Contaminant Properties

If one had perfect knowledge of the foreground contamination in the data cube, the problem of foreground contamination would be trivial; one would simply perform a direct subtraction of the foregrounds from the data vector $\mathbf{x}$. Unfortunately, our knowledge of foregrounds is far from perfect, particularly at the level of precision required for a direct detection of the cosmological 21 cm signal. Because of this, the estimator shown in Equation (4.6) in fact combines several different foreground subtraction steps in an attempt to achieve the lowest possible level of foreground contamination:

1. A direct subtraction of a foreground model from the data vector. This is given by $\mathbf{x} - \mathbf{m}$. To see this, note that the data vector can be thought of as being comprised of the cosmological 21 cm signal $\mathbf{x}_{21}$, the foregrounds $\mathbf{x}_{\text{fg}}$, and the instrumental noise $\mathbf{n}$. On the other hand, the mean data vector

$$\mathbf{m} \equiv \langle \mathbf{x} \rangle = \langle \mathbf{x}_{21} \rangle + \langle \mathbf{x}_{\text{fg}} \rangle + \langle \mathbf{n} \rangle = \langle \mathbf{x}_{\text{fg}} \rangle. \tag{4.14}$$

   contains only the foreground contribution, because we are interested in the *fluctuations* of the 21 cm signal, so the cosmological signal has zero mean, as does the instrumental noise (in the absence of instrumental systematics). Note that because the mean here is the mean in the ensemble average sense (as opposed to just the spatial mean), $\mathbf{m}$ represents a full spatial and spectral model of the foregrounds.

2. Since the foregrounds also appear in the covariance matrix, the action of $\mathbf{C}^{-1}$ is to downweight foreground-contaminated modes, exploiting foreground properties such as smooth frequency dependence.

3. Subtracting the term $\frac{1}{2}\text{tr}[\mathbf{C}_{\text{junk}}\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}]$ eliminates the bias from contaminants.

4. Finally, the binning of the cylindrical power spectrum to the spherical power spectrum provides yet more foreground suppression. Foregrounds are distributed in select regions on the $k_\perp$-$k_\parallel$ plane (i.e., outside the EoR window) in patterns that do not lie along contours of constant $k = \sqrt{k_\perp^2 + k_\parallel^2}$. Thus, when binning along such contours to produce a spherical power spectrum, one can selectively downweight parts of the contour with greater foreground contamination, which constitutes a form of foreground cleaning. Roughly speaking, this corresponds to taking advantage of the fact that foregrounds have a cylindrical symmetry in Fourier space, whereas the signal is spherically isotropic [153]. We do note, however, that the formalism we introduce in Section 4.2.7 is general enough to use any geometric differences between foregrounds and signal.

Of these foreground mitigation strategies, the first and third are direct subtractions (in amplitude and power, respectively), whereas the second and the fourth act through weightings. The former group represent operations that are particularly vulnerable to incorrectly modeled foregrounds. To see this, recall that the foregrounds are expected to be larger than the cosmological signal by three or four orders of magnitude [53, 107, 18, 182]. Thus, when performing direct subtractions, low-level, unaccounted-for inaccuracies in the foreground model can translate into extremely large biases in the final results. In addition, significant numerical errors may arise from the subtraction of two large numbers (the data and the foregrounds) to obtain a small number (the measured cosmological signal).

Our goal for the rest of the section is to immunize ourselves against biases from direct subtractions. Of the direct subtraction steps list above, the Step 1 is likely to be relatively harmless for two reasons. First, it is immediately followed by the $\mathbf{C}^{-1}$ downweighting. The downweighting mitigates the effects of inaccuracies in modeling, for the $\mathbf{C}^{-1}$ tends to gives less weight to precisely the modes that have the largest foreground amplitudes, and therefore would be the most susceptible to modeling errors in the first place. In addition, the uncertainty in foreground properties in those regions of the $k_\perp$-$k_\parallel$ plane result in large error bars there, providing a convenient marker of the untrustworthy parts of the plane, effectively demarcating the boundaries of the

EoR window. For these two reasons, Step 1 is unlikely to be an issue, at least not inside the EoR window.

More worrisome is Step 3, where the power spectrum bias of contaminants is subtracted off. If we define "contaminants" to be "everything but the cosmological 21 cm signal", there are two potential sources of bias: foregrounds and noise. The subtraction of these biases is not followed by a downweighting analogous to the application $\mathbf{C}^{-1}$ in Step 1. Moreover, whereas one could argue that the foreground bias is likely to be large only outside the EoR window, the noise bias will spread throughout the $k_{\perp}$-$k_{\parallel}$ plane. This noise bias will also be quite large, as current experiments are firmly in the regime where the signal-to-noise is below unity. It would therefore be advantageous to avoid bias subtractions altogether if possible.

To avoid having to subtract foreground bias, we simply redefine what we mean by contaminants/junk. If we modify our mission to be one where we are measuring the power spectrum of total sky emission instead of the power spectrum of the cosmological 21 cm signal, the foreground contribution to the bias term no longer exists, as foregrounds now count as part of the signal we wish to measure. Of course, nothing has really changed, for we have simply ignored the subtraction of the foreground bias by redefining what we mean by "contaminants". The method is still optimal for measuring the power spectrum of the sky emission—though now it will not provide the absolute best possible limits on the EoR power spectrum. Within the EoR window, this should result in little degradation of our final constraints, for in this region foreground contamination is expected to be negligible, and the power spectrum of the cosmological signal should be essentially identical to the power spectrum of total sky emission. In any case, this is an assumption that can be checked in the final results, and represents a conservative assumption throughout Fourier space since foreground power is necessarily positive. As detailed low-frequency foreground observations are conducted, it may be possible to achieve more sensitivity in foreground contaminated regions by taking advantage of more detailed maps and developing more faithful models. This task is left to future power spectrum estimation studies.

In contrast, escaping to the safe confines of the EoR window alone is not sufficient to eliminate the instrumental noise portion of the bias term, for the instrumental noise bias pervades the entire $k_\perp$-$k_\parallel$ plane. To eliminate the noise bias, one can choose to compute not the auto-power spectrum of a single data cube with itself, but instead to compute the cross-power spectrum of two data cubes that are formed from data from interleaved (i.e., odd and even) time samples. Since the instrumental noise is uncorrelated in time, this has the effect of automatically removing the instrumental noise bias[4].

More explicitly, we can form a bandpower estimate of the cross-power spectrum by simply computing

$$\widehat{p}_\alpha^{\text{cross}} = \mathbf{x}_1^t \mathbf{E}^\alpha \mathbf{x}_2, \tag{4.15}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the data vectors for the two time inter-leaved data cubes, and for notational brevity we have defined $\mathbf{E}^\alpha \equiv \frac{1}{2} \sum_\beta M_{\alpha\beta} \mathbf{C}^{-1} \mathbf{C}_{,\beta} \mathbf{C}^{-1}$. For notational cleanliness we will omit the $-\mathbf{m}$ term in our power spectrum estimator for this section only, with the understanding that $\mathbf{x}$ signifies the data vector *after* the best-guess foreground model has already been subtracted. In a similar fashion, $\mathbf{x}_{\text{fg}}$ refers to the foreground residuals, post-subtraction.

To see that the cross-power spectrum has no noise bias, let us decompose the data vectors $\mathbf{x}_i$ into the sum of $\mathbf{s}$ and $\mathbf{n}_i$, the signal and noise components respectively, where the signal component has no index because it does not vary in time (note also that following the discussion above, any true sky emission counts as signal, so that $\mathbf{s} \equiv \mathbf{x}_{21} + \mathbf{x}_{\text{fg}}$). Inserting this decomposition into the preceding equation and taking

---

[4]The reader may object to this by (correctly) pointing out that there exist errors that are correlated in time, with calibration errors being a prime example. The result would be a cross-power spectrum that still retained a bias. However, this does not invalidate the cross-power spectrum approach, in the following sense. While biases will make our estimates of the power spectrum imperfect, these estimate will not be incorrect—the final (biased) power spectra will still represent perfectly rigorous upper limits on the cosmological power, provided we are conservative about how we estimate our error bars. We will discuss how to make such conservative error estimates later on in this section and in Section 4.3.3.

the expectation value of the result gives

$$
\begin{aligned}
\langle \widehat{p}_\alpha^{\mathrm{cross}} \rangle &= \langle (\mathbf{s} + \mathbf{n}_1)^t \mathbf{E}^\alpha (\mathbf{s} + \mathbf{n}_2) \rangle \\
&= \langle \mathbf{s}^t \mathbf{E}^\alpha \mathbf{s} \rangle + \langle \mathbf{n}_1 \rangle^t \mathbf{E}^\alpha \mathbf{s} \mathbf{s}^t \mathbf{E}^\alpha \langle \mathbf{n}_2 \rangle + \langle \mathbf{n}_1 \mathbf{E}^\alpha \mathbf{n}_2 \rangle \\
&= \langle \mathbf{s}^t \mathbf{E}^\alpha \mathbf{s} \rangle,
\end{aligned}
\tag{4.16}
$$

where the last equality holds because the instrumental noise has zero mean, i.e. $\langle \mathbf{n}_i \rangle = 0$, and no cross-correlation between different times, i.e. $\langle \mathbf{n}_1 \mathbf{n}_2 \rangle = 0$. The resulting estimator depends only on the power spectrum of the signal, and there is no additive bias.

Importantly, however, we emphasize that while we have eliminated noise *bias* by computing a cross-power spectrum, we have not eliminated noise *variance*. In other words, the instrumental nosie will still contribute to the error bars. To see this, consider the variance in our estimator, which is given by

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\alpha\beta}^{\mathrm{cross}} &= \langle \widehat{p}_\alpha^{\mathrm{cross}} \widehat{p}_\beta^{\mathrm{cross}} \rangle - \langle \widehat{p}_\alpha^{\mathrm{cross}} \rangle \langle \widehat{p}_\beta^{\mathrm{cross}} \rangle \\
&= \langle \mathbf{x}_1^t \mathbf{E}^\alpha \mathbf{x}_2 \mathbf{x}_1^t \mathbf{E}^\beta \mathbf{x}_2 \rangle - \langle \mathbf{x}_1^t \mathbf{E}^\alpha \mathbf{x}_2 \rangle \langle \mathbf{x}_1^t \mathbf{E}^\beta \mathbf{x}_2 \rangle
\end{aligned}
\tag{4.17}
$$

The second term simplifies to

$$
\langle \widehat{p}_\alpha^{\mathrm{cross}} \rangle \langle \widehat{p}_\beta^{\mathrm{cross}} \rangle = \sum_{ijkl} \langle \mathbf{x}_1^i \mathbf{x}_2^j \rangle \langle \mathbf{x}_1^k \mathbf{x}_2^l \rangle \mathbf{E}_{ij}^\alpha \mathbf{E}_{kl}^\beta.
\tag{4.18}
$$

Similarly, the first term is equal to

$$
\begin{aligned}
\langle \widehat{p}_\alpha^{\mathrm{cross}} \widehat{p}_\beta^{\mathrm{cross}} \rangle &= \sum_{ijkl} \langle \mathbf{x}_1^i \mathbf{x}_2^j \mathbf{x}_1^k \mathbf{x}_2^l \rangle \mathbf{E}_{ij}^\alpha \mathbf{E}_{kl}^\beta \\
&= \sum_{ijkl} \left(\langle \mathbf{x}_1^i \mathbf{x}_2^j \rangle \langle \mathbf{x}_1^k \mathbf{x}_2^l \rangle + \langle \mathbf{x}_1^i \mathbf{x}_1^k \rangle \langle \mathbf{x}_2^j \mathbf{x}_2^l \rangle + \langle \mathbf{x}_1^i \mathbf{x}_2^l \rangle \langle \mathbf{x}_1^k \mathbf{x}_2^j \rangle \right) \mathbf{E}_{ij}^\alpha \mathbf{E}_{kl}^\beta,
\end{aligned}
\tag{4.19}
$$

where in the last equality we assumed Gaussian distributed data to simplify the four-point correlation.[5] Our bandpower covariance is now

$$\mathbf{\Sigma}_{\alpha\beta}^{\text{cross}} = \sum_{ijkl} \left(\langle \mathbf{x}_1^i \mathbf{x}_1^k \rangle \langle \mathbf{x}_2^j \mathbf{x}_2^l \rangle + \langle \mathbf{x}_1^i \mathbf{x}_2^l \rangle \langle \mathbf{x}_1^j \mathbf{x}_2^k \rangle \right) \mathbf{E}_{ij}^{\alpha} \mathbf{E}_{kl}^{\beta}. \tag{4.20}$$

The first term in this expression consists only of auto-correlations, which contain both noise and signal:

$$\langle \mathbf{x}_1 \mathbf{x}_1^t \rangle = \langle (\mathbf{s} + \mathbf{n}_1)(\mathbf{s}^t + \mathbf{n}_1^t) \rangle - \langle \mathbf{s} \rangle \langle \mathbf{s} \rangle^t = \mathbf{S} + \mathbf{N} = \mathbf{C}, \tag{4.21}$$

where we have defined $\mathbf{C}$ to be the total data covariance (as defined in Section 4.2.1), $\mathbf{S} \equiv \langle \mathbf{s} \mathbf{s}^t \rangle - \langle \mathbf{s} \rangle \langle \mathbf{s} \rangle^t$ is the sky signal covariance (as per the discussion earlier in this section), and $\mathbf{N} \equiv \langle \mathbf{n}_1 \mathbf{n}_1^t \rangle = \langle \mathbf{n}_2 \mathbf{n}_2^t \rangle$ is the instrumental noise covariance. We have assumed that there is no correlation[6] between the sky emission and the instrumental noise, so that $\langle \mathbf{s} \mathbf{n}_1^t \rangle = \langle \mathbf{s} \mathbf{n}_2^t \rangle = 0$.

The second term in our bandpower covariance consists only of cross-correlations, and thus contains no noise covariance:

$$\langle \mathbf{x}_1 \mathbf{x}_2^t \rangle = \langle (\mathbf{s} + \mathbf{n}_1)(\mathbf{s}^t + \mathbf{n}_2^t) \rangle = \mathbf{S}. \tag{4.22}$$

Putting everything together, we obtain

$$\mathbf{\Sigma}_{\alpha\beta}^{\text{cross}} = \text{tr}\left[\mathbf{C}\mathbf{E}^{\alpha}\mathbf{C}\mathbf{E}^{\beta}\right] + \text{tr}\left[\mathbf{S}\mathbf{E}^{\alpha}\mathbf{S}\mathbf{E}^{\beta}\right]. \tag{4.23}$$

---

[5]In principle, $\mathbf{x}$ may exhibit departures from Gaussianity, since foregrounds are typically not Gaussian-distributed. However, there are several reasons to expect deviations from non-Gaussianity to be unimportant. First, the most flagrantly non-Gaussian foregrounds are typically those that are bright. When we analyze real data in Section 4.3, we alleviate this problem by analyzing only a relatively clean part of the sky. In addition, recall that in this section, $\mathbf{x}$ represents the data *after* a best-guess model of foregrounds has been subtracted from the original measurements. Thus, the crucial probability distribution to consider is not the foregrounds themselves, but rather the *deviations* from the foregrounds, which are likely to be better-approximated by a Gaussian distribution.

[6]Note that this assumption has nothing to do with whether or not the instrument is sky-noise dominated. A sky-noise dominated instrument will have instrumental noise whose *amplitude* depends on the sky temperature, but the actual noise fluctuations will still be uncorrelated with the sky signal.

This, then, is the error covariance of our cross power spectrum estimator. It gives less variance than the expression for the auto power spectrum, which in the notation of this section takes the form

$$\mathbf{\Sigma}_{\alpha\beta}^{\text{auto}} = 2\text{tr}\big[\mathbf{C}\mathbf{E}^{\alpha}\mathbf{C}\mathbf{E}^{\beta}\big]. \tag{4.24}$$

Despite this difference between equations 4.23 and 4.24, one may conservatively opt to use the above covariance matrix for the auto-power spectrum to estimate error bars even when using Equation (4.15) to estimate the power spectrum itself. In fact, it may be prudent to make this choice because there exists the possibility that the noise between interleaved time samples may not be truly uncorrelated, making the true errors closer to those described by $\mathbf{\Sigma}^{\text{auto}}$. In our worked example with MWA data in Section 4.3, we will conservatively use Equation (4.24) to estimate the errors of our cross-power spectrum. The task of characterizing the noise properties of the instrument thoroughly enough to eliminate this assumption is left to future work on a larger data set.

In summary, uncertainties in noise and foreground properties make it desirable to avoid trying to extract weak signals by performing subtractions between two large numbers (the contamination-dominated data and the possibly inaccurate contaminant models). Mathematically, the greatest concern comes with the subtraction of the noise and foreground biases from power spectra estimates. To deal with the residual noise bias, one may evaluate cross-power spectra between interleaved time samples rather than auto-power spectra. To deal with the foreground bias, one can conservatively elect to simply leave it in when placing upper limits on the cosmological signal, and rely on the robustness of the EoR window to separate out the foregrounds from the cosmological 21 cm signal. In effect, one can practice foreground avoidance rather than foreground subtraction, since the former (if it is sufficient for a detection of the cosmological signal) will be more robust than the latter in the face of foreground uncertainties. Finally, as a brute-force safeguard, to quantify such uncertainties, one can always vary the foreground model used in power spectrum estimation, as we do

in Section 4.3.3 when we apply our methods to the worked example of MWA data.

4.2.4 A Real-World Obstacle: Incomplete uv-Coverage

While the methods of the previous section allow one to alleviate the effects of foreground modeling uncertainty, it is impossible to avoid the fact that real interferometers are imperfect imaging instruments. This is because a real interferometer will inevitably have uv-coverage that is non-ideal in two ways. First, the coverage is non-uniform, resulting in images that have been convolved with non-trivial synthesized beam kernels. Second, the uv-coverage is incomplete, in that certain parts of the uv-plane are not sampled at all. The idealized methods of Section 4.2.1 deals with neither problem, and in this section with augment the formalism to rectify this.

Assume for a moment that uv coverage is complete (so that there are no "holes" in the uv-plane), but not necessarily uniform. In such a scenario, one has measured an unevenly weighted sample of the Fourier modes of the sky. The effect of this non-trivial weighting needs to be accounted for when measuring the power spectrum, since uv coordinates roughly map to $k_\perp$. A failure to do so would therefore result in the final power spectrum estimate being multiplied by some function of $k_\perp$ corresponding to the uv distribution.

Put another way, the uv distribution of an interferometer defines its synthesized beam, the kernel with which the true sky has been convolved in the production of our image data cube. The equations of Section 4.2.1 assume that this convolution has already been undone. Thus, we must first perform this step, which in our notation may be written as

$$\mathbf{x} = \mathbf{B}^{-1}\mathbf{x}', \tag{4.25}$$

where $\mathbf{x}'$ represents the convolved data vector, $\mathbf{B}$ is the convolution matrix encoding the effects of the synthesized beam, and $\mathbf{x}$ is the processed data vector that is fed into Equation (4.6). Note that this application of $\mathbf{B}^{-1}$ is meant to undo only the effects of the synthesized beam, not the primary beam.

The above method assumes that the matrix $\mathbf{B}$ is invertible. In practice, this will

likely not be the case as parts of the uv plane will be missed by the interferometer, resulting in a singular $\mathbf{B}$ matrix. In what follows, we will present two different ways to deal with this. The first is to modify the equations of Section 4.2.1 so that they accept the convolved images (the "dirty maps") as input. Since all the statistical information relevant to the power spectrum are encoded in the covariance matrix, we simply have to make the replacement

$$\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^t \rangle - \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^t \longrightarrow \langle \mathbf{x}'\mathbf{x}'^t \rangle - \langle \mathbf{x}' \rangle \langle \mathbf{x}' \rangle^t. \tag{4.26}$$

This amounts to

$$\mathbf{C} \longrightarrow \mathbf{B} \left(\langle \mathbf{x}\mathbf{x}^t \rangle - \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^t \right) \mathbf{B}^t = \mathbf{B}\mathbf{C}\mathbf{B}^t. \tag{4.27}$$

Of course, changing the covariance matrix also changes $\mathbf{C}_{,\alpha}$, and we must propagate this change. Differentiating the preceding equation with respect to the bandpower p_α gives the substitution

$$\mathbf{C}_{,\alpha} \longrightarrow \mathbf{B}\mathbf{C}_{,\alpha}\mathbf{B}^t. \tag{4.28}$$

Since $\mathbf{C}_{,\alpha}$ is the response of the data covariance $\mathbf{C}$ to the bandpower p_α, this is simply a statement of the fact that if our data consists of dirty maps, the revised $\mathbf{C}_{,\alpha}$ matrix should encode the response of a dirty map's data covariance to the bandpower. With the substitutions given by Equations (4.27) and (4.28), the rest of the equations of Section 4.2.1 can be used unchanged. In the limit of an invertible $\mathbf{B}$ matrix, it is straightforward to show that this is equivalent to using Equation (4.25).

The second method for dealing with a singular $\mathbf{B}$, which was proposed in Ref. [58], is to replace the ill-defined inverse matrix $\mathbf{B}^{-1}$ with a pseudoinverse given by

$$\mathbf{\Pi} \left(\mathbf{B} + \gamma \mathbf{U}\mathbf{U}^\dagger \right)^{-1} \mathbf{\Pi}, \tag{4.29}$$

where γ is a non-zero but otherwise arbitrary real number, and $\mathbf{\Pi}$ is a projection matrix given by

$$\mathbf{\Pi} \equiv \mathbf{I} - \mathbf{U}(\mathbf{U}^\dagger\mathbf{U})^{-1}\mathbf{U}^\dagger. \tag{4.30}$$

The matrix $\mathbf{U}$ specifies which modes on the sky are missing in the data as a result of unobserved pixels on the uv-plane. It is constructed by computing the responses (on the sky) of each unobserved uv pixel individually and storing each response as a column of $\mathbf{U}$. As an example, in the flat-sky approximation the $\mathbf{U}$ matrix would have a sinusoid in each column, corresponding to the fringes that would have been observed by the interferometer had data not been missing in a particular uv pixel. If these modes were present in the covariance model (which might be the case, for example, if the covariance were constructed by modeling data from a different interferometer with different uv coverage), then the inverse covariance $\mathbf{C}^{-1}$ in our estimator needs to be similarly replaced with the pseudoinverse:

$$\mathbf{\Pi}\left(\mathbf{C} + \gamma\mathbf{U}\mathbf{U}^{\dagger}\right)^{-1}\mathbf{\Pi}. \tag{4.31}$$

Importantly, the pseudoinverse can be quickly multiplied by a vector using the previously discussed conjugate gradient method. Its usage therefore does not sacrifice any of the speedups that were identified in Section 4.2.2 for dealing with large data volumes.

4.2.5   A Real-World Obstacle: Missing Data from RFI

In any practical observation, the presence of narrowband RFI will mean that certain RFI-contaminated frequency channels will need to be flagged as outliers and omitted from a final power spectrum analysis. The result, once again, is the presence of gaps in the data, only this time the missing modes are complete frequency channels. However, the pseudoinverse formalism of the previous section is quite flexible in that modes of any form can be projected out of the analysis. Thus, to correctly account for RFI-flagged data, one simply uses the pseudoinverse in exactly the same way as one does to account for missing uv data.

4.2.6 A Real-World Obstacle: Foreground Leakage into the EoR Window

As Equation (4.11) showed, estimates of the power spectrum are not truly local, in the sense that every bandpower estimate $\widehat{p}_\alpha$ corresponds to a weighted average of the true power spectrum, with weights specified by the window functions. Liu and Tegmark [120] showed that these window functions can be quite broad, particularly in regions with high foreground contamination. There is thus the danger that foreground power could leak into the EoR window. Because the foregrounds are so much brighter than the cosmological signal, even a small amount of leakage could compromise the cleanliness of the EoR window.

Fortunately, one can exert some control over the shape of the window functions[7] by making wise choices regarding the form of $\mathbf{M}$ in Equation (4.8), which in turn gives the window functions via $\mathbf{W} = \mathbf{MF}$. As discussed above, $\mathbf{M}$ must be chosen such that the rows of $\mathbf{W}$ sum to unity. Beyond that requirement, however, an infinite number of choices are admissible. One choice would be $\mathbf{M} = \mathbf{F}^{-1}$, which gives $\mathbf{W} = \mathbf{I}$ (i.e., delta function windows). This would certainly minimize the amount of leakage into the EoR window, but it comes at a high price: the resulting error bars on the power spectrum measurement—the diagonal elements of $\mathbf{\Sigma}$ from Equation (4.9)—tend to be large, reflecting the data's inability to make highly localized measurements in Fourier space when the survey volume is finite.

On the other extreme, the error bars predicted by $\mathbf{\Sigma}$ can be shown to be their smallest possible if $\mathbf{M}$ is taken to be diagonal [214]. However, this gives broader window functions, for it is via the smoothing/binning effect of these broad window functions that the small errors can be achieved. One can also argue that the level of smoothing dictated by this approach is excessive, since the resulting bandpowers have positively correlated errors. (To see this, note that up to a row-dependent normalization, the error covariance matrix takes the form $\mathbf{\Sigma} \sim \mathbf{F}$. Since all elements of a

---

[7]The term "window function" should not be confused with the term "EoR window". The former refers to the weights that specify the linear combination of the true bandpowers that each bandpower estimate represents, as per Equation (4.11). The latter refers to the region on the $k_\perp$-$k_\parallel$ plane that naturally has very low levels of foreground contamination, as illustrated in Figure 4-1.

Fisher matrix must necessarily be non-negative, this implies that all cross-covariances of the estimated bandpowers have positively correlated errors unless $\mathbf{F}$ is diagonal, which is rarely the case).

As a compromise option, we advise $\mathbf{M} \sim \mathbf{F}^{-1/2}$ (again after a normalization of each row so that the window functions sum to unity). This choice gives window functions that are narrower than those for a diagonal $\mathbf{M}$ while maintaining reasonably small error bars. In addition, an inspection of Equation (4.9) reveals that this method gives a diagonal $\boldsymbol{\Sigma}$, which means that errors between different bandpowers are uncorrelated.

In Section 4.3.4, we use MWA data to demonstrate the crucial role that the $\mathbf{M} \sim \mathbf{F}^{-1/2}$ choice plays in preserving the cleanliness of the EoR window.[8]

4.2.7 A Real-World Obstacle: Ensuring that Binning doesn't Destroy Error Properties

In previous sections, we have discussed how one can preserve all the desirable properties of the power spectrum estimator of Section 4.2.1 in the face of all the real-world complications presented in Sections 4.2.2 through 4.2.6. The result is a rigorous yet practical estimator for the cylindrical power spectrum $P_{\text{cyl}}(k_\perp, k_\parallel)$. We now turn to the problem of binning the cylindrical power spectrum into the cosmologically relevant spherical power spectrum $P_{\text{sph}}(k)$, with a special emphasis on the preservation of the information content of our estimator.

Just as with the cylindrical power spectrum, we parameterize the spherical power spectrum as piecewise constant, so that all the information is encoded in a vector of bandpowers $\mathbf{p}^{\text{sph}}$, so that:

$$p_\alpha^{\text{sph}} \equiv P_{\text{sph}}(k^\alpha). \tag{4.32}$$

The spherical bandpowers are related to estimates of the cylindrical bandpowers $\widehat{\mathbf{p}}^{\text{cyl}}$

---

[8]Of course, there exist other choices that are more elaborate than the three considered in this paper. For example, with exquisite foreground and instrumental modeling, one could imagine first decorrelating to delta-function windows by setting $\mathbf{M} = \mathbf{F}^{-1}$ in an attempt to "perfectly" contain the foregrounds to regions outside the EoR window, and then to re-smooth the bandpowers within the window to reduce the variance. This is a promising avenue for future investigation, but for this paper our goal is simply to apply the $\mathbf{F}^{-1/2}$ decorrelator to real data (see Section 4.3.4) to demonstrate the feasibility of containing foregrounds using decorrelation techniques.

by the equation

$$\widehat{\mathbf{p}}^{\text{cyl}} = \mathbf{A}\mathbf{p}^{\text{sph}} + \boldsymbol{\varepsilon}, \tag{4.33}$$

where $\mathbf{A}$ is a matrix of size $N_{\text{cyl}} \times N_{\text{sph}}$ of 1s and 0s that relates $k_\perp$-$k_\parallel$ pairs to k bins, with N_{cyl} and N_{sph} equal to the number of cells in the $k_\perp$-$k_\parallel$ plane and the number of spherical k bins respectively. The vector $\boldsymbol{\varepsilon}$ is a random vector of errors on $\widehat{\mathbf{p}}^{\text{cyl}}$. It has zero mean (assuming that one has taken the care to avoid additive bias in our estimator of the cylindrical bandpowers, as discussed above), but non-zero covariance equal to $\boldsymbol{\Sigma}^{\text{cyl}} \equiv \langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t \rangle$, where $\boldsymbol{\Sigma}^{\text{cyl}}$ is given by either Equation (4.23) or (4.24), depending on whether the cylindrical bandpowers were computed using cross or auto-power spectra. (The methods presented in this section are applicable either way).

Our goal is to construct an optimal, unbiased estimator of $\mathbf{p}^{\text{sph}}$ from $\widehat{\mathbf{p}}^{\text{cyl}}$. This is a solved problem [209], and the best estimator $\widehat{\mathbf{p}}^{\text{sph}}$ is given by

$$\widehat{\mathbf{p}}^{\text{sph}} = [\mathbf{A}^t\boldsymbol{\Sigma}_{\text{cyl}}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\boldsymbol{\Sigma}_{\text{cyl}}^{-1}\widehat{\mathbf{p}}^{\text{cyl}}, \tag{4.34}$$

with the final error covariance on the spherical bandpowers given by

$$\boldsymbol{\Sigma}_{\alpha\beta}^{\text{sph}} \equiv \langle \widehat{\mathbf{p}}_\alpha^{\text{sph}}\widehat{\mathbf{p}}_\beta^{\text{sph}} \rangle - \langle \widehat{\mathbf{p}}_\alpha^{\text{sph}} \rangle \langle \widehat{\mathbf{p}}_\beta^{\text{sph}} \rangle = [\mathbf{A}^t\boldsymbol{\Sigma}_{\text{cyl}}^{-1}\mathbf{A}]^{-1}. \tag{4.35}$$

Since the $\mathbf{A}$ matrix has (by construction) a single 1 per row and zeros everywhere else, an inspection of Equation (4.35) reveals that a diagonal $\boldsymbol{\Sigma}^{\text{cyl}}$ implies a diagonal $\boldsymbol{\Sigma}^{\text{sph}}$. In other words, the estimator given by Equation (4.34) preserves the decorrelated nature of the $\mathbf{M} \sim \mathbf{F}^{-1/2}$ version of the cylindrical power spectrum estimator defined in Section 4.2.6. This will not be the case for an arbitrary estimator (such as one that is formed from taking uniformly weighted Fast Fourier Transforms, then squaring and binning). We also emphasize that if one does not choose to use decorrelated cylindrical bandpower vectors, Equations (4.34) and (4.35) require that one keep full track of the off-diagonal terms of $\boldsymbol{\Sigma}_{\text{cyl}}^{-1}$. Without it, a consistent propagation of errors to the spherical power spectrum is not possible, and may even lead to a mistakenly

claimed detection of the cosmological signal, as we discuss in Section 4.3.4 and in Appendix 4.A.

Just as with the cylindrical power spectra, we would like to compute the window functions. The definition of the spherical window functions are exactly analogous to that provided in Equation (4.11) for the cylindrical power spectrum, so that

$$\langle \widehat{\mathbf{p}}^{\text{sph}} \rangle = \mathbf{W}^{\text{sph}} \mathbf{p}^{\text{sph}}. \tag{4.36}$$

Taking the expectation value of Equation (4.34), we have

$$\begin{aligned} \langle \widehat{\mathbf{p}}^{\text{sph}} \rangle &= [\mathbf{A}^t \boldsymbol{\Sigma}_{\text{cyl}}^{-1} \mathbf{A}]^{-1} \mathbf{A}^t \boldsymbol{\Sigma}_{\text{cyl}}^{-1} \langle \widehat{\mathbf{p}}^{\text{cyl}} \rangle \\ &= [\mathbf{A}^t \boldsymbol{\Sigma}_{\text{cyl}}^{-1} \mathbf{A}]^{-1} \mathbf{A}^t \boldsymbol{\Sigma}_{\text{cyl}}^{-1} \mathbf{W}^{\text{cyl}} \mathbf{A} \mathbf{p}^{\text{sph}}, \end{aligned} \tag{4.37}$$

where we have used the definition of the cylindrical window functions to say that $\langle \widehat{\mathbf{p}}^{\text{cyl}} \rangle = \mathbf{W} \mathbf{p}^{\text{cyl}}$, as well as the fact that $\mathbf{p}^{\text{cyl}} = \mathbf{A} \mathbf{p}^{\text{sph}}$ (with no error term because we are relating the true cylindrical bandpowers to the true spherical bandpowers). Inspecting this equation, we see that

$$\mathbf{W}^{\text{sph}} = [\mathbf{A}^t \boldsymbol{\Sigma}_{\text{cyl}}^{-1} \mathbf{A}]^{-1} \mathbf{A}^t \boldsymbol{\Sigma}_{\text{cyl}}^{-1} \mathbf{W}^{\text{cyl}} \mathbf{A}. \tag{4.38}$$

Therefore, by measuring the width of the spherical window functions (rows of $\mathbf{W}^{\text{sph}}$), one can place rigorous horizontal error bars on the final spherical power spectrum estimate.

4.2.8 Summary of the issues

In the last few sections, we have provided techniques for dealing with a number of real-world obstacles. These include:

1. Taking advantage of the flat-sky approximation and the rectilinearity of data cubes, as well as the conjugate gradient algorithm for matrix inversion to allow large data sets to be analyzed quickly.

2. Using cross-power spectra rather than auto-power spectra in order to eliminate noise bias.

3. Replacing inverses with pseudoinverses to deal with data that has missing spatial modes (due to incomplete *uv* coverage) and missing frequency channels (due to RFI).

4. Performing power spectrum decorrelation to avoid the leakage of foreground power into the EoR window.

5. Binning of cylindrical power spectra into spherical power spectra in a way that preserves desirable error properties.

Crucial to this is the fact that these techniques all operate under a self-consistent framework. This allows faithful error propagation that accurately captures how various real-world effects act together. For example, it was shown in [58] that properly accounting for pixelization effects in Equation (4.5) results in low Fisher information at high $k_\parallel$, providing a marker for parts of the $k_\perp$-$k_\parallel$ plane that cannot be well-constrained because of finite spectral resolution. The identification of such a region would be trivial if one had spectrally contiguous data, for then one would simply say that the largest measurable $k_\parallel$ was roughly $1/\Delta L_\parallel$, where $\Delta L_\parallel$ is the width of a single frequency channel mapped into a cosmological line-of-sight distance. However, such a straightforward analysis no longer applies when there are RFI gaps in the data at arbitrary locations. In contrast, the unified framework presented in this paper allows all such complications to be folded in correctly.

4.3   A Worked Example: Early MWA Data

Now that we have bridged the gap between theoretical techniques for analyzing ideal data and the numerous challenges presented by real data, we are ready to bring together our methods, specify a covariance model, and estimate power spectra from MWA 32-tile prototype (MWA-32T) data. The data were taken between the 21st and

Field of View (Primary Beam Width)	$\sim 25°$ at 150 MHz
Angular Resolution	$\sim 20'$ at 150 MHz
Collecting Area	~ 690 m^2 towards zenith at 150 MHz
Polarization	Linear X-Y
Frequency Range	80 MHz to 300 MHz
Instantaneous Bandwidth	30.72 MHz
Spectral Resolution	40 kHz

Table 4.1: MWA-32 Instrument Parameters

29th of March 2010, the first observing campaign during which data were taken that were scientifically useful. The observations are described in more detail by [234]. Real data affords us two opportunities. In this section, we look at the data to examine and quantify the differences between power spectrum estimators and the pitfalls associated with choice of estimator. In Section 4.4, we take advantage of everything we have developed to arrive at interesting new foreground results and a limit on the 21 cm brightness temperature power spectrum.

4.3.1 Description of Observations

All of the data used for this paper were taken on the MWA-32T system. This system has since been upgraded to a 128-tile instrument (MWA-128T; Tingay et al. [220], Bowman et al. [29]), but in this paper we focus exclusively on MWA-32T data, reserving the MWA-128T data for future work.

The MWA-32T instrument consisted of 32 phased-array "antenna tiles" which served as the primary collecting elements. Each tile contained 16 dual linear-polarization wideband dipole antennas which were combined to form a steerable beam with a full width at half maximum (FWHM) size of $\sim 25°$ at 150 MHz. The array had an approximately circular layout with a maximum baseline length of ~ 340 m, and a minimum baseline length of 6.6 m, although the shortest operating baseline during this observational campaign was 16 m. After digitization, filtering, and correlation, the final visibilities had a 1 second time resolution and 40 kHz spectral resolution over a 30.72 MHz bandwidth. The instrumental capabilities are summarized in Table 4.1.

For our worked example, we concentrate on March 2010 observations of the MWA "EoR2" field. It is centered located at R.A.(J2000) = 10^{h} 20^{m} 0^{s}, decl.(J2000) = $-10°$ $0'$ $0''$, and is one of two fields at high Galactic latitude that have been identified by the MWA collaboration as candidates for deep integrations, owing to their low brightness temperature in low frequency measurements of Galactic emission [88, 53]. For further details about the observational campaign or the EoR2 field, please see Williams et al. [234], which was based on the same set of observations as the ones used in this paper.

Observations covered three 30.72 MHz wide bands, centered at 123.52 MHz, 154.24 MHz and 184.96 MHz, corresponding to a redshift range of $6.1 < z < 12.1$ (the redshift range of the results presented in this work is slightly smaller because of data flagging) for the 21 cm signal. The 123.52 MHz and 154.24 MHz bands were observed for approximately 5 hours each, and the 184.96 MHz band was observed for approximately 12 hours.

These early data from the prototype have provided us with a set of test data that enabled development of extensive analysis methods and software on which the results of this paper are based. The early prototype had shortcomings (e.g., mismatched cables, receiver firmware errors, correlator timing errors) that compromised the calibration to some extent, raising the apparent noise level. Additionally, the instrument was only operating with $\lesssim 29$ tiles, and with a 50% duty cycle throughout the course of these observations. We account for this in Section 4.3.3 by determining the magnitude of the noise empirically, in order to be able to place rigorously conservative upper limits on the cosmological power spectrum. We expect that data from later prototype campaigns and from the full array will produce result closer to theoretical expectations.

4.3.2   Mapmaking

Before the data can be used as a worked example for our power spectrum estimator, however, we must convert the measured visibilities into a data cube of sky images at every frequency in our band. In other words, we must form the data vector $\mathbf{x}$, defined

231

by Equation (4.5), which serves as the input for our power spectrum pipeline.

To form the data vector, we performed the following steps. First, we performed a reduction procedure similar to that described in Williams et al. [234] for the initial flagging and calibration of the data. Hydra A was identified as the dominant bright source in the field, and used for calibration assuming a point source model. The Hydra A source model was then subtracted from the uv data. As this same source model was also used for gain and phase calibration, this can be thought of as a "peeling" source removal procedure [159, 228, 149, 100] on a single source. Alternatively, in the absence of gridding artifacts, this is equivalent to imaging the point-source model and subtracting it from the data as part of the direct foreground subtraction step discussed in the first step of Section 4.2.3 [209].

The subtracted data were imaged using the CASA task `clean` without deconvolution to produce "dirty" images. No multi-frequency synthesis was performed, so that the full 40 kHz spectral resolution of the data would be available. The visibilities were gridded using w-projection kernels [47] with natural (inverse-variance) weighting to produce maps at each frequency with a cell size of $3'$ over a $25.6°$ field of view. The resulting cubes contained ~ 200 million voxels, with 512 elements along each spatial dimension and 768 elements in the frequency domain. It is important to note that the pre-flagging performed on the data resulted in the flagging of entire frequency bands (which means that there are gaps in the final data cube). Cubes were generated for each 5 minute snapshot image.

The individual snapshot data cubes were combined using the primary beam inverse-variance weighting method described in Williams et al. [234]. The weighting and primary beams were simulated separately for each 40 kHz frequency channel in each 5 minute snapshot. The combined maps and weights were saved, along with the effective point spread function at the center of the field. Two additional data cubes were created by averaging alternating 5 minute snapshots (i.e. even numbered snapshots were averaged into one cube, and odd numbered snapshots were averaged into the other) so that they were generated from independent data, but with essentially the same sky and uv coverage properties.

A further flux scale calibration of the integrated cubes was performed using three bright point sources: MRC 1002-215, PG 1048-090, and PKS 1028-09 to set the flux scale on a channel-by-channel basis. A two dimensional Gaussian fitting procedure was used to fit the peak flux of each of these sources in each 40 kHz channel of the data cube. Predictions for each source were derived by fitting a power law to source measurements from the 4.85 GHz Parkes-MIT-NRAO survey [83], the 408 MHz Molonglo Reference Catalog [114], the 365 MHz Texas Survey [62], the 160 MHz and 80 MHz Culgoora Source List [200] and the 74 MHz VLA Low-frequency Sky Survey [45]. A weighted least-squares fit was then performed to calculate and apply a frequency-dependent flux scaling for the cube to minimize the square deviations of the source measurements from the power law models.

An additional flagging of spectral channels was performed based on the root-mean-square (RMS) noise in each spectral channel of the cube. A smooth noise model was determined by median filtering the RMS channel noise as a function of frequency (bins of 16 channels were used in the filtering). Any channel with 5σ or larger deviations from the smoothed noise model was flagged. Upon inspection, these additional flagged channels were observed to be primarily located at the edges of the coarse digital filterbank channels, which were corrupted due to an error in the receiver firmware. After this procedure, approximately one third of the spectral channels were found to have been flagged.

Each individual map covered $25.6° \times 25.6°$ at a resolution of $3'$ with 768 frequency channels (40 kHz frequency resolution). To decrease the computational burden of the covariance estimation, each map was subdivided into 9 subfields, and the pixels were averaged to a size of $15'$. The data cubes were mapped to comoving cosmological coordinates using WMAP-7 derived cosmological parameters, with $\Omega_{\mathrm{M}} = 0.266$, $\Omega_{\Lambda} = 0.734$, $H_0 = 71$ km s^{-1} Mpc^{-1}, and $\Omega_k \equiv 0$ [111].

At this point, the data cubes were ready to be used as input data to our power spectrum estimator, i.e., we had arrived at the final form of the data vector $\mathbf{x}$. However, estimating power spectra and error statistics using the formalism of Section 4.2 also requires a covariance model, which we construct in the next section.

4.3.3 Covariance Model

We follow [120] and [58] in modeling the covariance matrix $\mathbf{C}$ as the sum of independent parts attributable to noise and foregrounds. We leave off the signal covariance because it only contributes to the final error bars by accounting for cosmic variance—a completely negligible effect in comparison to foreground and noise-induced errors. We adopt a conservative model of the extragalactic foregrounds by treating them as a Poisson random field of sources with fluxes less than 100 Jy, after the manual removal of Hydra A. By treating all extragalactic foregrounds as "unresolved," we effectively throw out information about which lines of sight are most contaminated by bright foregrounds. As [58] showed, future analyses can improve on our limits by including more information about the foregrounds. We begin with the parameterized covariance model of [120],

$$
C_{ij}^{\text{unresolved}} = \left(1.4 \times 10^{-3} \, \frac{\text{K}}{\text{Jy}} \right)^2 \left(\int_0^{S_{\text{cut}}} S^2 \frac{dn}{dS} dS \right) \left(\frac{\nu_i \nu_j}{\nu_*^2} \right)^{-2-\bar{\kappa}} \times
$$

$$
\exp\left[\frac{\sigma_{\bar{\kappa}}^2}{2} \left(\ln\left[\frac{\nu_i \nu_j}{\nu_*^2} \right] \right)^2 \right] \times \exp\left[\frac{(\mathbf{r}_{\perp i} - \mathbf{r}_{\perp j})^2}{2\sigma_\perp^2} \right] \left(\Omega_{\text{pix}} \right)^{-1} \tag{4.39}
$$

where $\nu_* = 150 \, \text{MHz}$ is a reference frequency, ν_i is the frequency of the ith voxel, which has an angular distance of $\mathbf{r}_{\perp i}$ from the field center. The spectral index is $\bar{\kappa} = 0.5$, the uncertainty in the spectral index is $\sigma_\kappa = 0.5$, the clustering correlation length is $\sigma_\perp = 7'$, Ω_{pix} is the angular size of each pixel, the flux cut $S_{\text{cut}} = 100$ Jy, and dn/dS is the differential source count from [57],

$$
\frac{dn}{dS} = (4000 \, \text{Jy}^{-1}\text{sr}^{-1}) \times
\begin{cases}
\left(\frac{S}{0.880 \, \text{Jy}} \right)^{-2.51} & \text{for S} > 0.880 \, \text{Jy} \\
\left(\frac{S}{0.880 \, \text{Jy}} \right)^{-1.75} & \text{for S} \leq 0.880 \, \text{Jy.}
\end{cases}
$$

We adapt this model for the fast power spectrum estimation method outlined in Section 4.2.2 by calculating the translationally invariant approximation to this model in the manner described in [58].

For the Galactic synchrotron, we also follow [120] and [58] for the parameterization

234

of the synchrotron emission covariance. Namely, we adopt $\bar{\kappa} = 0.8$, $\sigma_\kappa = 0.1$, $\sigma_\perp = 30°$, and replace the first three terms of the covariance in Equation 4.39 with $T_{\mathrm{synch}}^2 = (335.4 \ \mathrm{K})^2$.

Our model for the instrumental noise is adopted from [58], with one key difference: the overall normalization. For each subband, we let the noise covariance matrix scale by a free multiplicative constant. This is equivalent to treating the combination $T_{\mathrm{sys}}^2/(A_{\mathrm{ant}}^2 t_{\mathrm{obs}})$ as a free parameter. We then fit for that parameter by requiring the RMS difference between the two time slices—which should be free of sky signal—for the densely sampled inner region of uv space and rescaling our noise covariance matrix to match. The spatial structure of the covariance was left unchanged. Even though the data is somewhat nosier than suggested by a first principles calculation assuming fiducial values for system temperature and antenna effective area, this empirical renormalization allows for an honest account of the errors introduced by instrumental effects.

To verify that our parameterization of the foregrounds was reasonable, we varied these parameters over an order of magnitude and found that they had little effect on our final power spectrum estimates, except at the lowest values of k. There are two reasons for this: first, since we are only measuring the power spectrum of the sky, we need not worry about precisely subtracting foregrounds. Second, because the noise in our instrument is still more than two orders of magnitude from the cosmological signal, in the EoR window our band power measurements will be noise dominated and agnostic to our foreground model. Future analyses might include a more thorough treatment of the foregrounds, especially by utilizing the full power of the Dillon et al. [58] method to include information about the positions, fluxes, and spectral indices of individual point sources.

4.3.4   Evaluating Power Spectrum Estimator Choices

With both a data vector $\mathbf{x}$ and a covariance matrix $\mathbf{C}$ in hand, we can now apply the methods of Section 4.2 to estimate power spectra. In doing so, we deal with real-world obstacles using all of the techniques that we have developed. In this section, we show

why all this is necessary.

In Section 4.2.6 we touted the choice of power spectrum estimator $\widehat{\mathbf{p}} = \mathbf{Mq}$ with $\mathbf{M} \sim \mathbf{F}^{-1/2}$ as a compromise solution between the choice with the smallest error bars, $\mathbf{M} \sim \mathbf{I}$, and the choice with the narrowest window functions, $\mathbf{M} \sim \mathbf{F}^{-1}$. In the race to detect the power spectrum from the EoR, one might be tempted to aggressively seek out the smallest possible errors. This could prove a deleterious choice, as we will now show using MWA-32T data.

First, in Figure 4-2 we compare cylindrical power spectra, $\widehat{\mathbf{p}}$, generated using two different estimators of the power spectrum that we presented in Section 4.2.6.[9] On the left, we have used $\mathbf{M} \sim \mathbf{I}$, the estimator with the smallest error bars, and on the right we have used $\mathbf{M} \sim \mathbf{F}^{-1/2}$, the estimator with decorrelated errors. In both cases, we have plotted the absolute value of the power spectrum estimates (which can be negative because they are cross-power spectra). Because the two estimates are related to one another by an invertible matrix, they contain the same cosmological information. In a sense, the $\mathbf{M} \sim \mathbf{F}^{-1/2}$ method is the most honest estimator of the power spectrum because the band powers form a mutually exclusive and collectively exhaustive set of measurements. In other words, they represent all the all the power spectrum information from the data, divided into independent pieces.

Moreover, just because two sets of estimators have the same information content does not mean that they are equally useful for distinguishing the cosmological power spectrum from foreground contamination. In Figure 4-2, the minimum variance estimator for the power spectrum introduces considerable foreground contamination into the EoR window, demarcated by the expected angular extent of the wedge feature (which we introduced in Section 4.1 and will discuss in greater detail in Section 4.4.1). Even highly suspect features at high $k_\perp$ where uv coverage is spottiest seem to get smeared across $k_\perp$ and into the EoR window. We cannot simply cut out the wedge from our cylindrical-to-spherical binning and expect a clean measurement of

---

[9]In our comparison of choices for $\mathbf{M}$, we drop the $\mathbf{M} \sim \mathbf{F}^{-1}$, δ-function windows choice. In addition to proving the noisiest estimator, it suffers from strong anti-correlated errors. We adopt the perspective that the important comparison is between the "obvious" choice, the minimum variance $\mathbf{M} \sim \mathbf{I}$, and our preferred choice with decorrelated errors, $\mathbf{M} \sim \mathbf{F}^{-1/2}$.
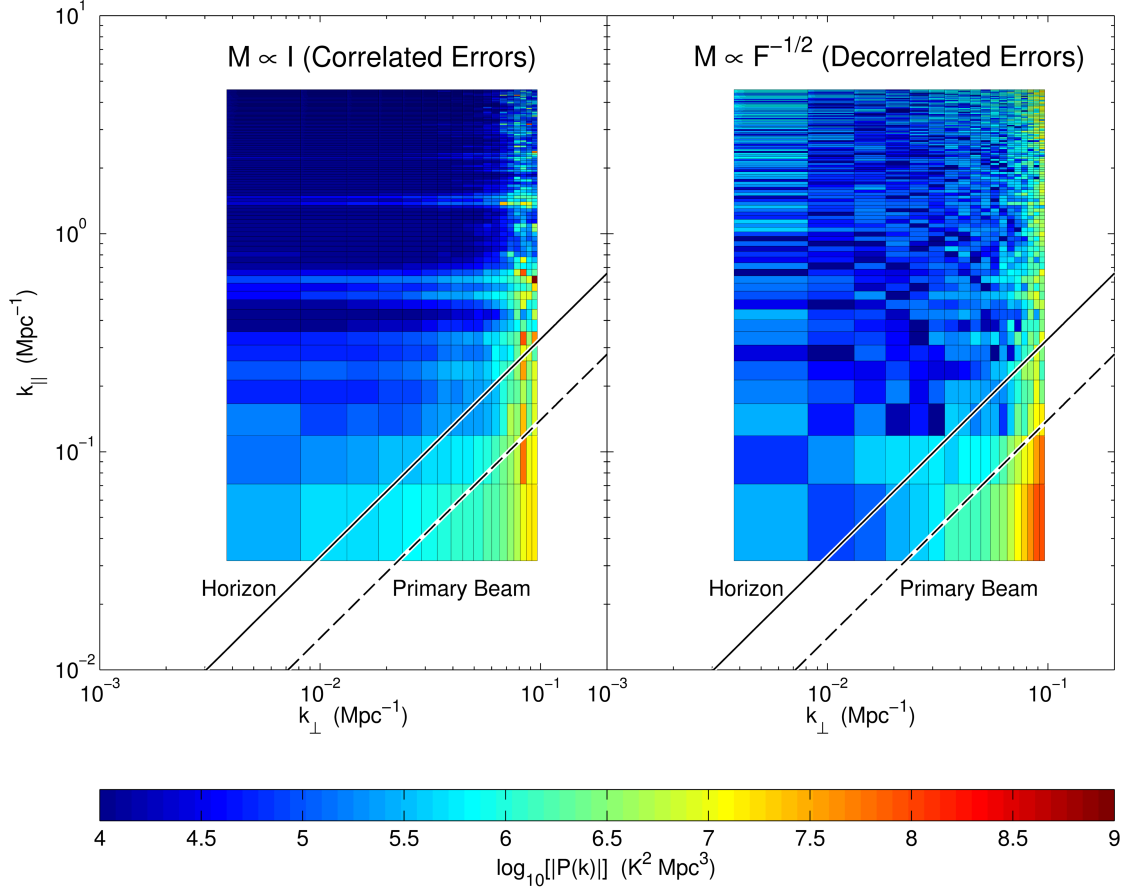
Figure 4-2: Unless one chooses a power spectrum estimator with decorrelated errors, foregrounds and other instrumental effects can leak significantly into the EoR window. Here we show the *absolute value* of the cylindrical power spectrum estimate from the subband centered on 158 MHz ($z = 8.0$) and averaged over all 9 fields. On the left, we have set $\mathbf{M} \sim \mathbf{I}$. On the right, $\mathbf{M} \sim \mathbf{F}^{-1/2}$. We expect contamination from smooth spectrum foregrounds interacting with the chromatic synthesized beam to occupy the "wedge" portion of Fourier space, defined in Equation (7.21). Optimistically, the wedge is delimited by the extent of the main lobe of the primary beam; conservatively, we should not see bright foreground contamination beyond the horizon. In the regions where the power spectrum is noise dominated, we expect little structure in the $k_{\parallel}$ direction in the EoR window above some moderate value of $k_{\parallel}$. In the left panel, we see considerably more $k_{\parallel}$ structure in the form of horizontal bands, attributable to foreground contamination and instrumental effects, that has leaked into the putative EoR window.

the power spectrum in the EoR window.

Looking closely at Figure 4-2, one might notice that some regions of the EoR window on the lefthand panel still seem very clean—cleaner perhaps that the same

regions in the righthand panel. To examine that apparent fact, we plot $\widehat{p}_\alpha$ instead of $|\widehat{p}_\alpha|$ in Figure 4-3. To make the figure more intelligible, we have plotted colors
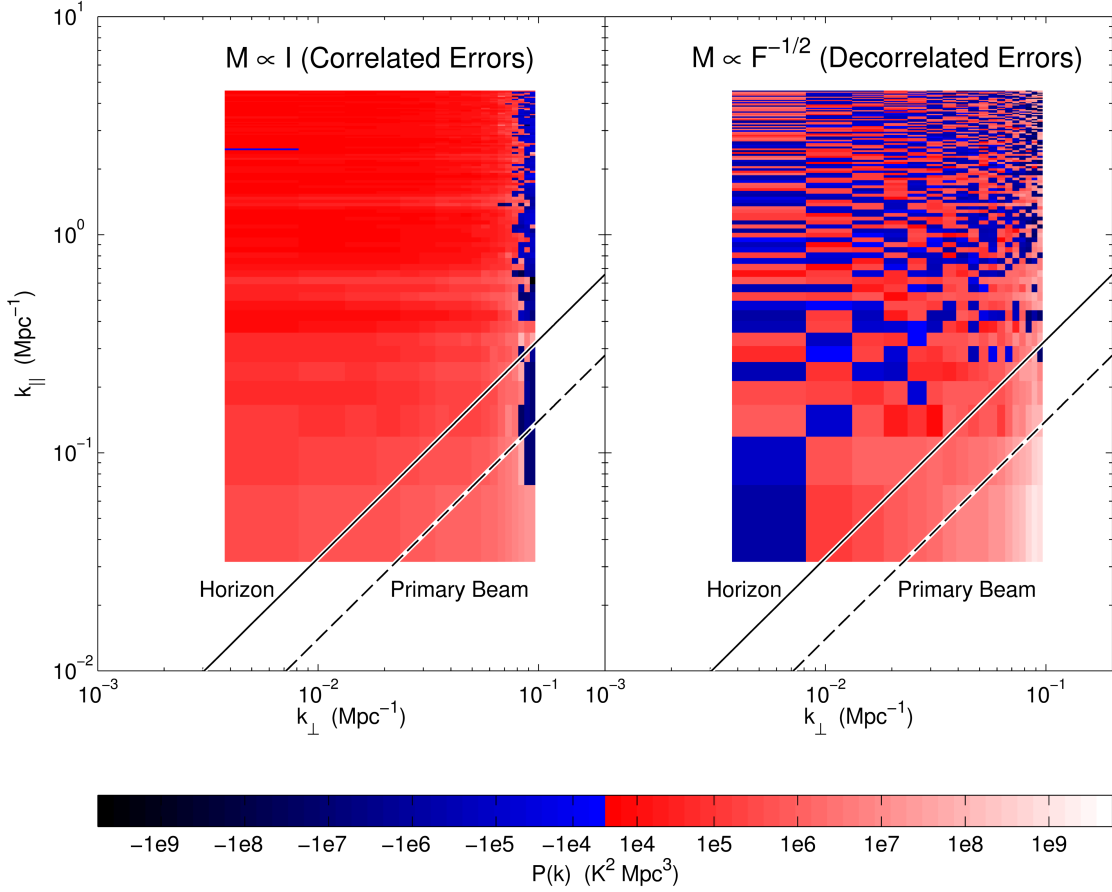
Figure 4-3: One advantage of calculating the cross power spectrum of interleaved time-slices of data is that we can easily tell which regions of Fourier space are noise dominated. Here we reproduce the power spectra from Figure 4-2 without taking the absolute value of $P(k)$. By plotting with a discontinuous, $\sinh^{-1}$ color scale, it is easy to see that the EoR window for our decorrelated power spectrum estimate (right panel) has roughly an equal number of positive and negative band power estimates—exactly what we would expect from a noise dominated region. By contrast, our power spectrum estimate with correlated errors (left panel) shows positive power over almost all of Fourier space, indicating ubiquitous leakage of contaminants into the EoR window.

based on an $\sinh^{-1}$ color scale with a sharp color division at 0. The $\sinh^{-1}$ has the advantage of behaving linearly at small values of $\widehat{p}_\alpha$ and logarithmically at large positive or negative $\widehat{p}_\alpha$.

What emerges is a striking difference between the two estimators. For the rea-

sons discussed in Section 4.2.3, we have chosen to estimate the cross power spectrum between two time-interleaved sets of observations. As a result, we expect that instrumental noise should be equally likely to contribute positive power as it is to contribute negative power. In noise dominated regions of the $k_\perp$-$k_\parallel$ plane, we expect about half of our measurements to be positive and about half to be negative. That is exactly what we see in the EoR window of the $\mathbf{M} \sim \mathbf{F}^{-1/2}$ estimator. However, the $\mathbf{M} \sim \mathbf{I}$ estimator in the lefthand panel clearly shows positive power throughout the entire supposed EoR window. Though the magnitude of that power is not enormous—often it is well within the vertical error bars—the overall bias towards positive cross power means that sky signal is contaminating the EoR window. This is precisely the problem we were worried about in Section 4.2.6 and the data have clearly manifested it.[10]

This also explains why there appeared to be less power in the EoR window of the lefthand panel of Figure 4-2; by taking the absolute value of the weighted average of positive and negative quantities, we expect to measure a smaller absolute value of the power. However, as this figure clearly shows, that weighted average is biased by foreground leakage. And, even though there still appears to be a region just inside the EoR window that retains positive band powers consistent with foregrounds, that small amount of leakage can be attributed to finite sized windows functions and to calibration uncertainties. Regardless, it does not appear to be an insurmountable limitation to the cleanliness of the EoR window; rather, it suggests that we should be careful in how we demarcate the EoR window when calculating spherically-averaged power spectra.

In addition to producing a cleaner EoR window, the decorrelated estimator of the power spectrum yields another advantage: narrower window functions. Both the estimator with the minimum variance and estimator with decorrelated errors repre-

---

[10]Of course, as we noted in Section 4.2.6, the choice of $\mathbf{M} \sim \mathbf{F}^{-1/2}$ is not unique in its ability to mitigate foreground leakage, and other choices certainly warrant future investigation. Picking $\mathbf{M} \sim \mathbf{F}^{-1/2}$ is, however, a good choice for a first attempt at decorrelation, particularly given its various other desirable properties that we have described. The important point here is that while $\mathbf{M} \sim \mathbf{F}^{-1/2}$ may not be necessarily optimal for containing foregrounds within the wedge, our results show that it is a reasonable one. In contrast, the "straightforward" approach of normalizing the power spectrum with the diagonal choice $\mathbf{M} \sim \mathbf{I}$ is clearly ill-advised.

sent, in aggregate, the weighted average of the true, underlying band power spectrum, as we discussed in Section 4.2.1. In Figure 4-4, we show the improvement that the decorrelated estimator offers over the minimum variance estimator by narrowing the window functions considerably.[11] We show five example window functions from the
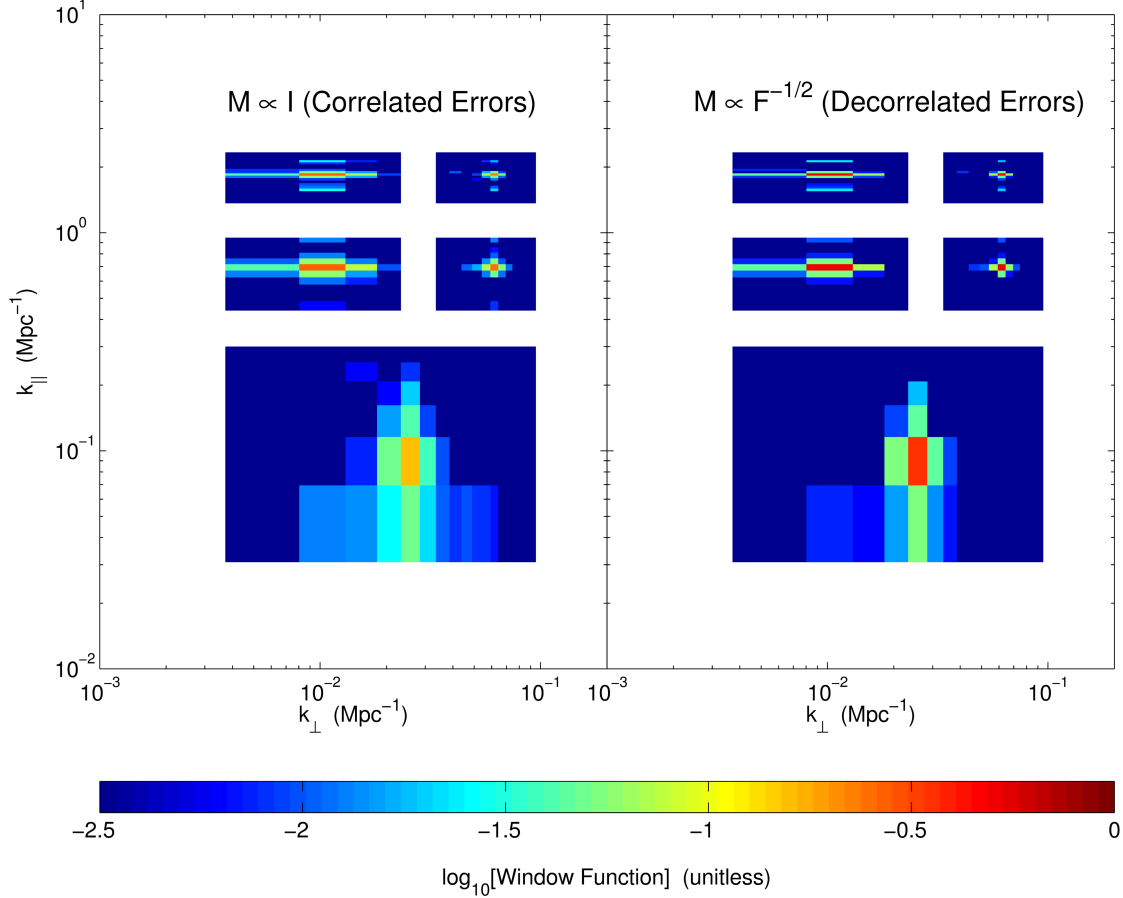
Figure 4-4: By using an estimator of the 21 cm power spectrum with uncorrelated errors, we significantly narrow the window functions that relate the ensemble average of our estimator to the true, underlying power spectrum. Here we show a sample of five cropped window functions for the power spectrum estimate in Figure 4-2, each centered at their maxima, for both an estimator with correlated errors (left panel) and an estimator with uncorrelated errors (right panel). Though the estimator with correlated errors produces smaller vertical error bars, it acheives this by "over-smoothing" many band powers together. Narrow window functions let us independently measure many modes of the power spectrum. The band power measured with $\mathbf{M} \sim \mathbf{F}^{-1/2}$ is one of a set of mutually exclusive and collectively exhaustive pieces of information.

---

[11]While the choice of $\mathbf{M} \sim \mathbf{F}^{-1/2}$ ensures that the power spectrum estimator covariance is diagonal (recall, $\boldsymbol{\Sigma} = \mathbf{M}\mathbf{F}\mathbf{M}^t$ while $\mathbf{W} = \mathbf{M}\mathbf{F}$), it does not mean that the window functions are delta functions. The off-diagonal terms of $\boldsymbol{\Sigma}$ might be zero even if the off-diagonal terms of $\mathbf{W}$ are not.

same subband that we plot in Figure 4-2, cropped to fit together on one set of axes, each centered at their respective peaks. Because the window functions are normalized to sum to 1, the breadth of each window function is reflected by the value of the central peak. As we expected, the window functions are considerably narrower for our decorrelated power spectrum estimator.

Even after binning from cylindrical power spectra to spherical power spectra, the difference remains quite stark. In Figure 4-5 we see clearly that choosing a power spectrum estimator with decorrelated errors also considerably improves the window functions in one dimension as well as two.

Lastly, as we mentioned in Section 4.2.7, one of advantage of our method is that it keeps a full accounting of the error covariance, $\mathbf{\Sigma}$. When $\mathbf{M}$ is not chosen to make $\mathbf{\Sigma}$ diagonal, an improper accounting can lead to a suboptimal or simply incorrect propagation of errors. In Appendix 4.A we work through an example of the consequences of assuming the independence of errors at various steps in the analysis. This should serve as a warning of the importance of careful analysis; incorrectly assuming a diagonal $\mathbf{\Sigma}$ can lead to unnecessarily wide window functions, an overestimation of errors, or—worst of all—an underestimation of errors that could lead to an unjustified claim of a detection.

4.4   Early Results

Having developed and demonstrated a technique that robustly preserves the EoR window while thoroughly and honestly keeping track of the errors on and correlations between our band power estimates, we can now confidently generate some interesting preliminary science results. Because these data span the widest redshift range to date, we are able to investigate the behavior of the wedge feature over many frequencies. Understanding the behavior of the EoR window over a large redshift range is important, since there is still considerable uncertainty about the timing and duration of the EoR. Moreover, it is often argued that a tentative first detection of the cosmological signal will only be convincingly distinguishable from residual foregrounds if one can
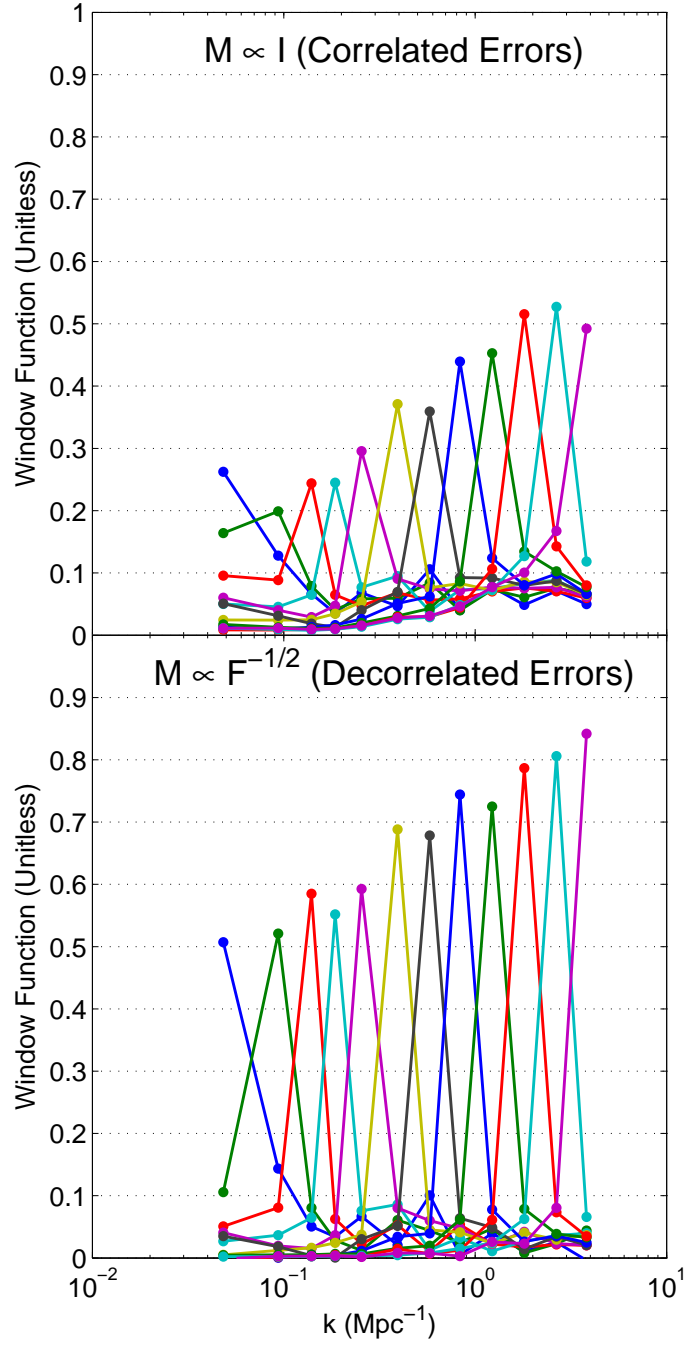
Figure 4-5: Even after optimally binning the cylindrical power spectra from Figure 4-2 to spherical power spectra, the choice of a power spectrum estimator with decorrelated errors produces much narrower window functions than the minimum variance technque. In addition to maintaining a clean EoR window, the choice of $\mathbf{M} \sim \mathbf{F}^{-1/2}$ provides the additional benefit of allowing power spectrum modes to be measured more independently.

show that the 21 cm brightness temperature fluctuations peak at some redshift, since theory predicts that the midpoint of reionization should be marked by such a peak [117, 22]. It is therefore essential to characterize the EoR window (and by extension, residual foregrounds) over a broad frequency range. We also apply our methods from Section 4.2 to calculate spherically averaged power spectra over our entire redshift range, including error bars and window functions, thus setting a limit on the 21 cm brightness temperature power spectrum during the EoR.

4.4.1 The Wedge

In Figure 4-6, we show all the cylindrical power spectra over the redshift range probed by our current observations. The spectra are sorted into three rows, each of which contain data coming from a single 30.72 MHz wide frequency band. All of the spectra were generated using the same techniques that were used to generate the example cylindrical power spectra in Section 4.3.4 and thus contain all the desirable statistical properties discussed in Section 4.2. One sees that in every case the foregrounds are mostly confined to the wedge region in the bottom right corner of the $k_\perp$-$k_\parallel$ plane. This builds upon the single frequency observations of [182], demonstrating the existence of the EoR window across a wide range of frequencies relevant to EoR observations.

Having these measurements also allows us to examine the behavior of the EoR window as a function of frequency. Consider first the high $k_\perp$ regions of the $k_\perp$-$k_\parallel$ plane. The most striking feature here is the wedge. Consistent with being dominated by foreground power, the wedge generally gets brighter with decreasing frequency within each wide frequency band, just as foreground emission is known to behave. The extent of the wedge is also in line with theoretical expectations. Recall from Equation (7.21) that the wider the field-of-view, the farther up in $k_\parallel$ the wedge goes. Since the field-of-view is defined by the primary beam, whose extent decreases with increasing frequency, one expects the wedge to have the largest area at the lowest frequencies. This trend is clearly visible in the cylindrical power spectra of Figure 4-6, where the wedge extends to the highest $k_\parallel$ at the highest redshifts. Importantly, the

wedge is confined to its expected location across the entire range of the observations. To see this, note that we have overlaid Equation (7.21) on the plots, with the dashed line corresponding to θ_{max} equal to that of the first null of the primary beam, and the solid line with $\theta_{\mathrm{max}} = \pi/2$ (the horizon). At all frequencies, the most serious contaminations lie within the first null, ensuring that the EoR window is foreground-free.

Foregrounds also enter indirectly into the instrumental noise-dominated regions because the MWA is sky-noise dominated. Thus, as the brightest sources of emission in our observations, the foregrounds set the system temperature, and result in a higher instrumental noise at higher redshifts. This trend can be seen within each wide frequency band (each row of Figure 4-6), although the slight interruption of this trend between bands suggests an additional source of noise.

At low $k_\perp$, theory suggests that foregrounds will contaminate a horizontally-oriented region at the bottom of the plot. This is clearly seen in the highest frequency plots. Interestingly, at lower frequencies the increasing instrumental noise plays more of a role, and the foreground contribution is less obvious in comparison (although it is still there). While a naive reading of some of these low frequency plots (such as the one for $z = 9.1$) might suggest that the EoR window extends to the lowest $k_\parallel$, such a conclusion would be misguided. As we shall see in Section 4.4.2, these modes are likely dominated by foregrounds (and therefore do not integrate down with further integration unlike instrumental noise dominated modes). Moreover, the error statistics (which self-consistently include foreground errors in our formalism) suggest that low $k_\parallel$ modes are less useful for constraining theoretical models, and that the true EoR window does in fact lie at higher $k_\parallel$, as suggested by theory. Again, this highlights the importance of estimating power spectra in a framework that naturally contains a rigorous calculation of the errors involved.

4.4.2 Spherical Power Spectrum Limits

Having confirmed that the EoR window behaves as expected, we will now proceed to place constraints on the spherical power spectrum. In top panel of Figure 4-7 we

show the result of binning the $z = 10.3$ cylindrical power spectrum of Figure 4-6, using the optimal binning formulae presented in Section 4.2.7. In addition, for ease of interpretation, we elect to plot

$$\Delta(k) \equiv \sqrt{\frac{k^3}{2\pi^2} P(k)} \qquad (4.40)$$

(which simply has units of temperature) rather than $P(k)$ itself.

To quantify the errors in our spherical power spectrum estimate, we also bin the cylindrical power spectrum measurement covariances and window functions using the formulae of Section 4.2.7. The resulting window functions are shown in the bottom panel, and give an estimate of the horizontal error bars. Thinking of these window functions (which, recall, are normalized to integrate to unity) as probability distributions, the horizontal error bars shown in the top panel are demarcated by the 20th and 80th percentiles of the distribution. (This corresponds to the full-width-half-maximum in the event that the window functions are Gaussians). The vertical error bars were obtained by taking the square root of each diagonal element of the covariance matrix. Since the methods of Section 4.2.7 carefully preserved the diagonal nature of the bandpower covariance, each data point in Figure 4-7 represents a statistically independent measurement. This would not have been the case had we not employed the decorrelation technique of Section 4.2.6.

Immediately obvious from the plot is that there is a qualitative difference between the data points at low k and those at high k. In particular, the points at low k are detections of the sky power spectrum, whereas the points at high k are formal upper limits. This is not to say, of course, that the cosmological EoR signal has been detected at low k. Rather, recall from Section 4.2.3 that in an attempt to avoid having to make large bias subtractions, we elected to compute cross-power spectra of total sky emission rather than of the cosmological signal, with the expectation (largely confirmed in Section 4.4.1) that the intrinsic cleanliness of the EoR window would be sufficient to ensure a relatively foreground-free measurement at high $k_\parallel$. Now, our survey volume is such that we are sensitive almost exclusively to regions in Fourier

245

space where $k_\parallel \gg k_\perp$. When binning along contours of constant k in the cylindrical Fourier space, we have that $k \equiv \sqrt{k_\perp^2 + k_\parallel^2} \approx k_\parallel$, and therefore the low k points of Figure 4-7 map to low $k_\parallel$. The detections seen at low k thus reside outside the EoR window and are almost certainly detections of the foreground power spectrum.

Despite the fact that the low k modes are foreground dominated, they still constitute a formal upper limit on the cosmological power spectrum, since the foreground power spectrum is necessarily positive. In fact, our current, most competitive upper limit resides at the lowest k values. However, this is unlikely to continue to be the case as more data is taken with the MWA, for two reasons. First, as foreground-limited measurements, the data points at low k will not average down with further integration time. In addition, the error statistics in the region are not particularly encouraging. The window functions (and therefore the horizontal error bars) are seen to broaden towards lower k, reducing the ability of constraints at those k to place limits on theoretical models. (This is most easily seen by recalling that the window functions integrate to unity by construction, and thus the increase in their peak values towards higher k implies a broadening of the window functions). The broadening of the window functions is an expected consequence of foreground subtraction [120] and thus will likely continue to limit the usefulness of the low k regime unless future measurements can characterize foreground properties with exquisite precision.

In contrast, the points at high k do reside in the EoR window. The constraints here are limited by thermal noise, as we saw in Section 4.3.4. Bolstering this view is the fact that the data here are consistent with zero, as one expects for a noise-dominated cross-power spectrum. The limits here are given by the 2σ errors predicted by the Equation (4.35). As mentioned in Section 4.3.3, these errors are somewhat larger than what might be predicted by a theoretical sensitivity calculation. However, they are consistent with rough estimates of the errors obtained from a calculation of root-mean-square values from the images produced in Section 4.3.2. This suggests that the larger-than-expected errors are due to noisier-than-expected input maps, and not to any approximations made in the power spectrum estimation techniques presented in this paper. The data on which these results are based are from the very first

operation of the prototype array, and we expect better performance in later data. Encouragingly, we note also that as noise-dominated constraints, the measurements at high k will continue to improve with integration time.

In Figure 4-8, we show power spectrum limits across the entire frequency range of the MWA, along with some theoretical predictions generated using the models in [10]. At the lowest redshift, no theory curve is plotted because the model predicts that reionization is complete by then. This yet again underscores the importance of making measurements over a broad frequency range—with access to a wide range of redshifts, future detections of the cosmological signal can be distinguished from residual foregrounds by measuring null signals at redshifts where reionization is complete.

Each redshift bin of Figure 4-8 exhibits trends that are qualitatively similar to those discussed above for the $z = 10.3$ case. We see many apparent detections of correlations positive correlations between the two time-interleaved data cubes—more than can be attributed to foregrounds alone. As we saw with the cylindrically binned power spectrum in Figure 4-2, there is evidence of systematic and instrumental effects sending foreground power into the EoR window, leading to higher k detections and large differences between neighboring k bins. With as new an instrument as the MWA was at time of this observation, this issues are understandable. The exact physical origin of those systematics is beyond the scope of this paper, however they should serve as a reminder to stay vigilant for them in future datasets from a more battle-tested instrument. However, because we see no evidence of strong anti-correlations between data cubes, we expect that the extra power introduced by systematics into the EoR window only the effect of worsening the limits we can set.

Over all bands, our best limit is $\Delta(k) < 0.3\,\mathrm{K}$, occurring at $z = 9.5$ and $k = 0.046\,\mathrm{cMpc}^{-1}$. However, as remarked in Section 4.3.3, the lowest k bins can be rather sensitive to the covariance model, and if one excludes those bins, our best limit is $\Delta(k) < 2\,\mathrm{K}$, at $z = 9.5$ and $k = 0.134\,\mathrm{cMpc}^{-1}$. While our limits may not be quite as low as other existing limits in the literature [167, 173], they are the only limits on the EoR power spectrum that span a broad redshift range from $z = 6.2$ to $z = 11.7$.

Moreover, these statistically rigorous limits will likely improve with newer and more sensitive data from the MWA.

4.5   Conclusions

In this paper, we have accomplished three goals. First, we adapted $21\,\mathrm{cm}$ power spectrum estimation techniques from Liu and Tegmark [120] and Dillon et al. [58] with real-world obstacles in mind, so that they could be applied to real data. With early MWA data, our generalized formalism was then used to demonstrate the importance of employing a statistically rigorous framework for power spectrum estimation, lest one corrupt the naturally foreground-free region of Fourier space known as the EoR window. Finally, we used the MWA data to set limits on the EoR power spectrum.

In confronting real-world obstacles, our desire is to preserve the as much of the statistical rigor in previous matrix-based power spectrum estimation frameworks as possible. To avoid having to perform direct subtractions of instrumental noise biases, we advocate computing cross-power spectra between statistically identical subsets of the data (in the case of the MWA worked example of this paper, these subsets were formed from odd and even time samples of the data). This has the effect of eliminating noise bias in the power spectrum, although instrumental noise continues to contribute to the error bars. To avoid direct subtractions of foreground biases, we simply look preferentially in the EoR window, where foregrounds are expected to be low. Missing data, whether from incomplete uv coverage or RFI flagging, can be dealt with using the pseudoinverse formalism. Doing this allows the effects of missing data to be self-consistently propagated into error statistics such as the power spectrum covariance and the window functions. In an effort to preserve the cleanliness of the EoR window, one should form decorrelated bandpower estimates, which have uncorrelated errors and reasonably narrow window functions. Care must then be taken to preserve these nice properties via an optimal binning of cylindrical bandpowers into spherical bandpowers.

Using early MWA data to demonstrate these techniques, we have confirmed the-

oretical predictions for the existence of the EoR window and have extended previous observations done by other groups to a much wider frequency range. This allowed us to check predicted trends of the EoR window as a function of frequency, all of which are consistent with theory. Crucially, we found that without using the decorrelation technology of Section 4.2.6, measurements in the EoR window are not in fact instrumental noise dominated, and contain a systematic bias that is indicative of foreground leakage from outside the EoR window.

The early MWA data has also allowed us to place limits on the cosmological EoR power spectrum. Our best limit is $\Delta(k) < 0.3\,\mathrm{K}$, at $z = 9.5$ and $k = 0.046\,\mathrm{cMpc}$ (or $\Delta(k) < 2\,\mathrm{K}$ at $z = 9.5$ and $k = 0.134\,\mathrm{cMpc}^{-1}$ if one discards the lowest k bin to immunize oneself against foreground modeling uncertainties). This may not be competitive with other published observations, but generalizes them in an important way: instead of focusing on one particular frequency, our limits span a wide range of redshifts relevant to the EoR, going from $z = 6.2$ to $z = 11.7$. In addition, these limits will almost certainly improve in the near future, using already-collected (but yet to be analyzed) data from the MWA-32T system, as well as soon-to-be-collected data from the MWA-128T system. The rigorous statistical tools developed in this paper should be equally applicable to these newer data sets, ensuring that foreground contamination remains confined to outside the EoR window, safeguarding the potential of current generation experiments to make an exciting first detection of the EoR within the next few years.

4.A   Appendix: On the Importance of Modeling the Full Error Covariance

In Section 4.2.7, we argued that an inverse covariance weighted binning scheme for estimating spherical band powers produced optimal spherical power spectrum estimate. In the case where $\mathbf{M}$ is chosen either for the smallest possible error bars or the narrowest possible window functions, the estimator covariance $\boldsymbol{\Sigma}^{\mathrm{cyl}}$ is non-

diagonal. Assuming that the matrix is actually diagonal, at one or more steps in the binning and error propagation, can lead error bars that are overly conservative or—worse yet—error bars that are insufficiently conservative and might falsely lead to a claimed detection. In Figure 4-9, we show the effects of making a suboptimal choice for binning.

If one fully models the covariance matrix $\mathbf{\Sigma}^{\text{cyl}}$, including off-diagonal elements, but chooses to generate $\widehat{\mathbf{p}}^{\text{sph}}$ as an inverse variance (and not inverse covariance) weighted average of cylindrical band powers, neglecting off diagonal terms in the weighting, one's estimators will be noisier as a result (see the solid lines in Figure 4-9). These are the correct errors for the suboptimal choice of estimators.

Even worse, if one assumes that $\mathbf{\Sigma}^{\text{cyl}}$ is diagonal when it is not, one is led either to overestimate the error bars, in the case of $\mathbf{M} \sim \mathbf{F}^{-1}$, or underestimate them, as would be the case when $\mathbf{M} \sim \mathbf{I}$. This is because the former case general exhibits anti-correlated errors while the latter suffers from correlated errors. The last scenario is the most troubling: by aggressively choosing the estimator with the smallest vertical error bars ($\mathbf{M} \sim \mathbf{I}$) and then neglecting the correlations between errors, one will underestimate the error bars and might be lead to falsely claiming a detection. In this case, the estimator is suboptimal and the errors are incorrect. Additionally, as we show in Figure 4-10, if one were to calculate the the window functions under the assumption that $\mathbf{\Sigma}^{\text{cyl}}$ is diagonal, one would find window functions several times boarder than they would otherwise be.

Thankfully, choosing the cylindrical power spectrum estimator with decorrelated errors avoids the subtle difference between inverse variance and inverse covariance weighted binning. The $\mathbf{M} \sim \mathbf{F}^{-1/2}$ decorrelated estimator preserves the EoR window and allows for easy, optimal binning of uncontaminated regions into spherical band power spectrum estimates.
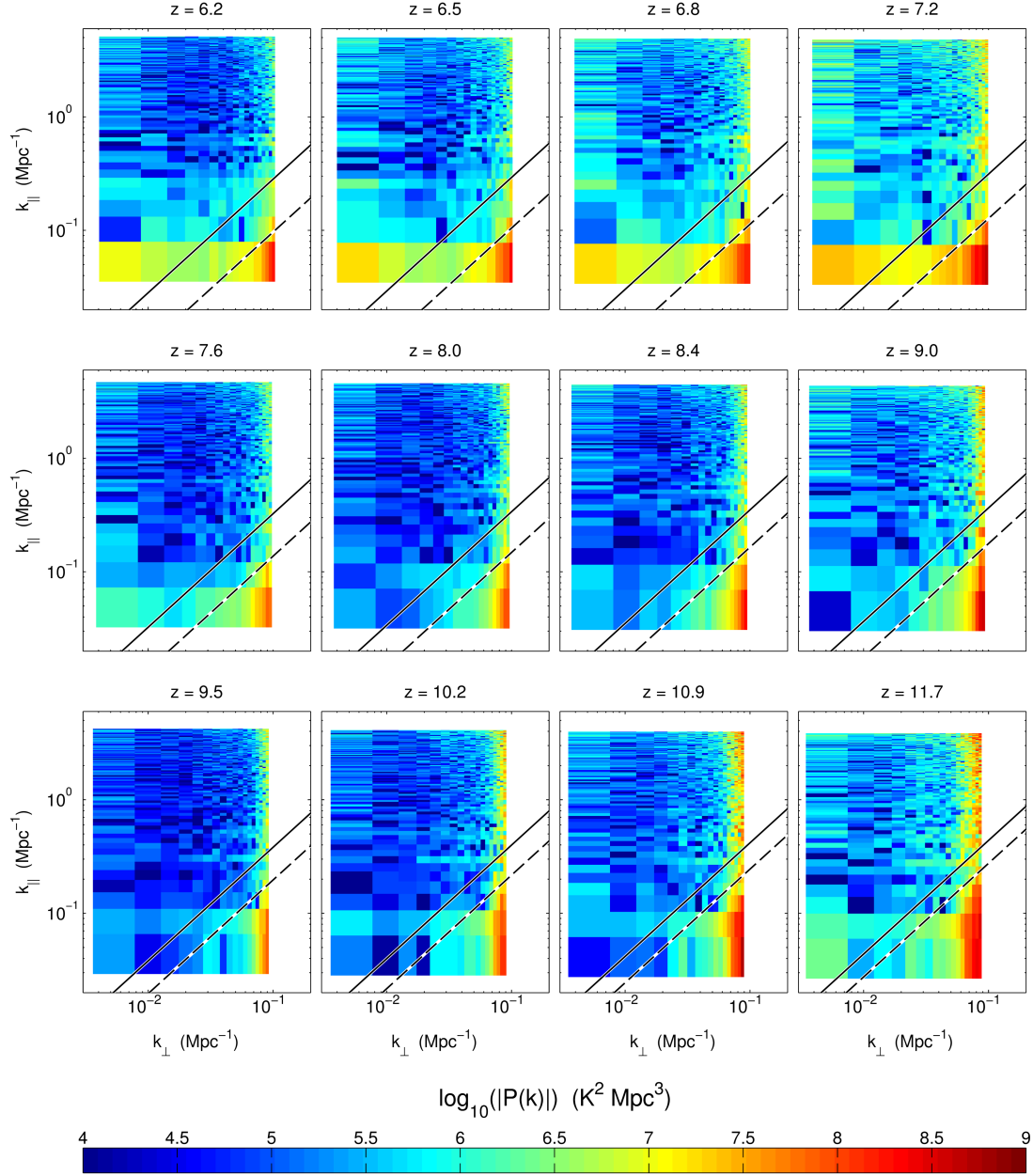
Figure 4-6: Examining cylindrically binned power spectra for each subband (each averaged over all nine subfields), reveals several important trends with frequency of the EoR window and the foregrounds. Each row is a single simultaneously observed frequency band. Since different bands were observed for different amounts of time, direct comparisons between rows is challenging. However, several clear trends emerge. For each band, moving to higher redshift (increasing wavelength) shows stronger foregrounds, a larger wedge (in part due to a wider primary beam), and a noisier EoR window (due to a higher system temperature). In general the brightest foreground contamination is well demarcated by the wedge line in Equation (7.21) for the primary beam (dotted line) and especially by the wedge line for the horizon (solid line). In short, the wedge displays the theoretically expected frequency dependence.
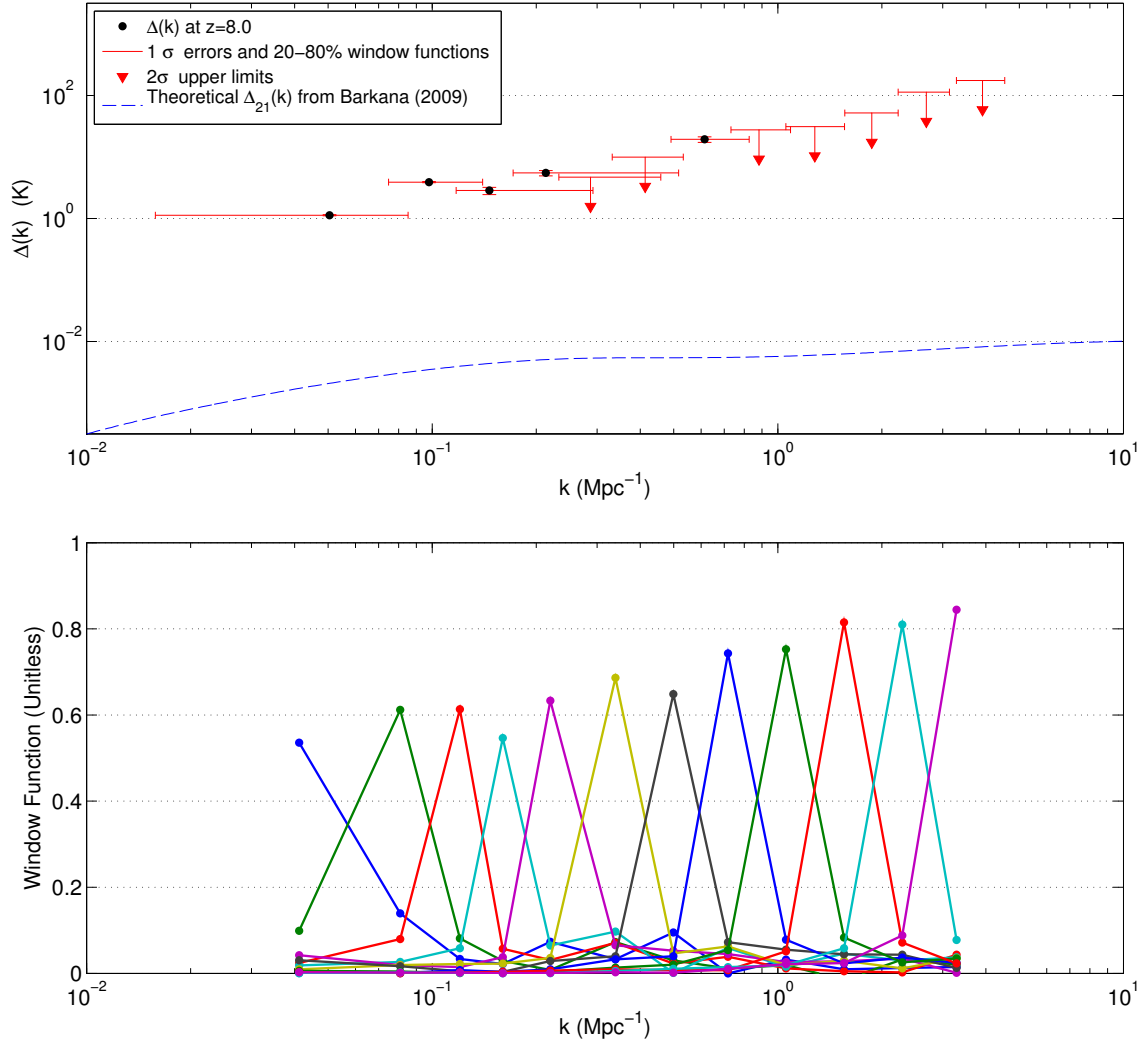
Figure 4-7: Our method allows for the estimation of the spherically binned power spectrum in temperature units, $\Delta(k)$, while keeping full acount of both vertical error bars and window functions (horizontal error bars) and making an optimal choice in the tradeoff between the two. In the top panel, we have plotted our spherical power spectrum estimates of the subband centered on 158 MHz ($z = 8.0$), including 1σ errors on detections (which are often only barely visible), 2σ upper limits on non-detections, and horizontal error bars that span the middle three quintiles of the window functions (bottom panel). At low k, the wide error bars are the expected consequence of foreground contamination [120]. Downward arrows represent measurements consistent with noise at the 2σ level. Even though the area under the primary beam wedge has been excised from the 2D-to-1D binning, the detection of foregrounds at low k, is expected due to the contribution of unresolved foregrounds over a wide range of $k_\perp$ [58]. Our fiducial theoretical power spectrum is taken from [10].
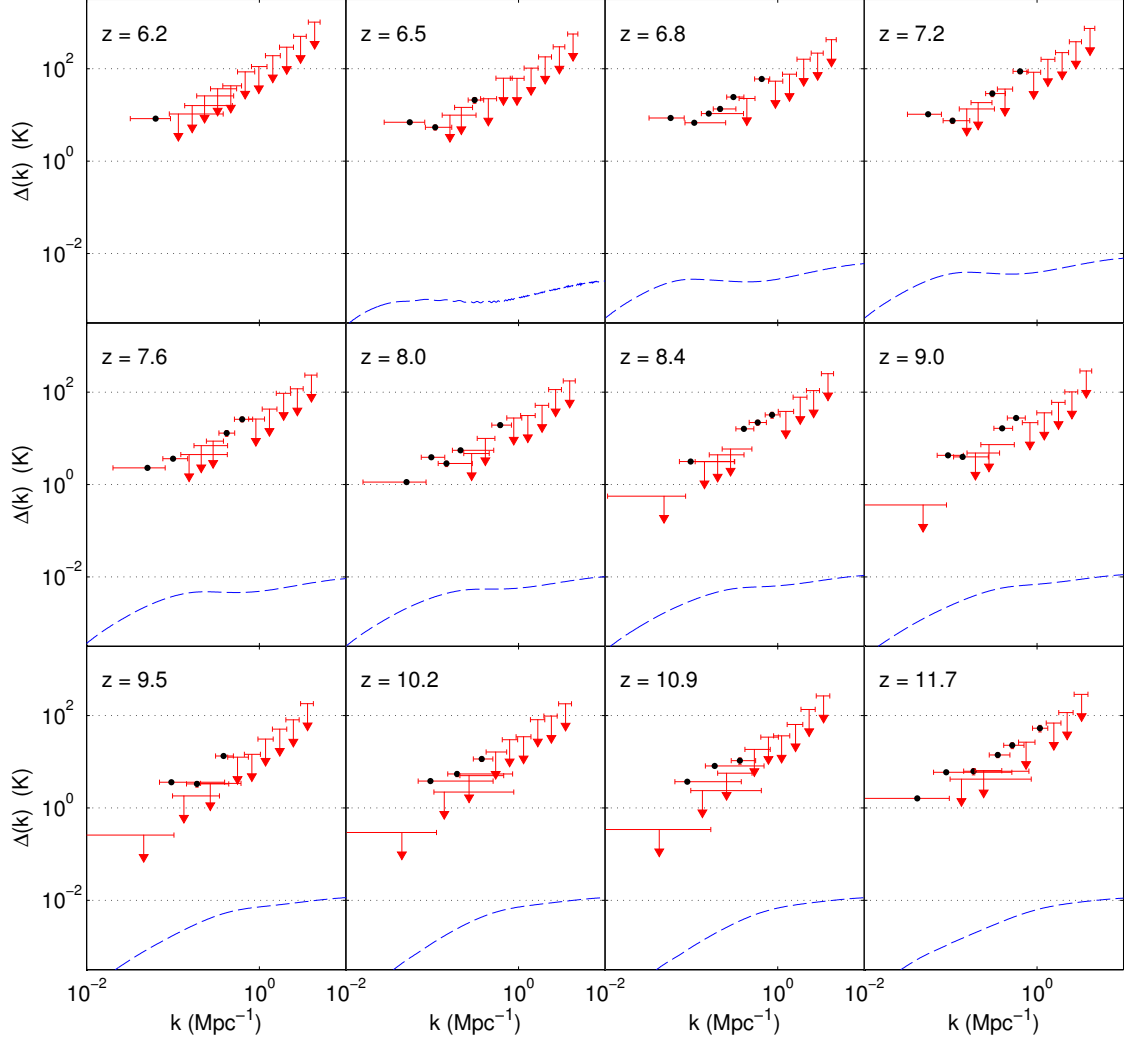
Figure 4-8: Taking advantage of our fast yet thorough power spectrum estimation technique, we estimate $\Delta(k)$ for a wide range of k and z, including both vertical and horizontal errors. (For points that represent positive detections of foregrounds or systematic correlations, the vertical error bars are often barely visible). Using the visual language of Figure 4-7, we show here our spherical power spectrum limits as a function of both k and z. Each panel is a different subband. The many detections can be attibuted to foregrounds (especially at low k), instrumental effects like those we saw in Figure 4-2 (especially at medium values of k), or both. Our absolute lowest limit on the 21 cm brightness temperature power spectrum, $\Delta(k) < 0.3$ Kelvin at the 95% confidence level, comes at $k = 0.046$ cMpc^{-1} and $z = 9.5$ (or $\Delta(k) < 2\,\mathrm{K}$ at $z = 9.5$ and $k = 0.134\,\mathrm{cMpc}^{-1}$ if one discards the lowest k bin to immunize oneself against foreground modeling uncertainties).
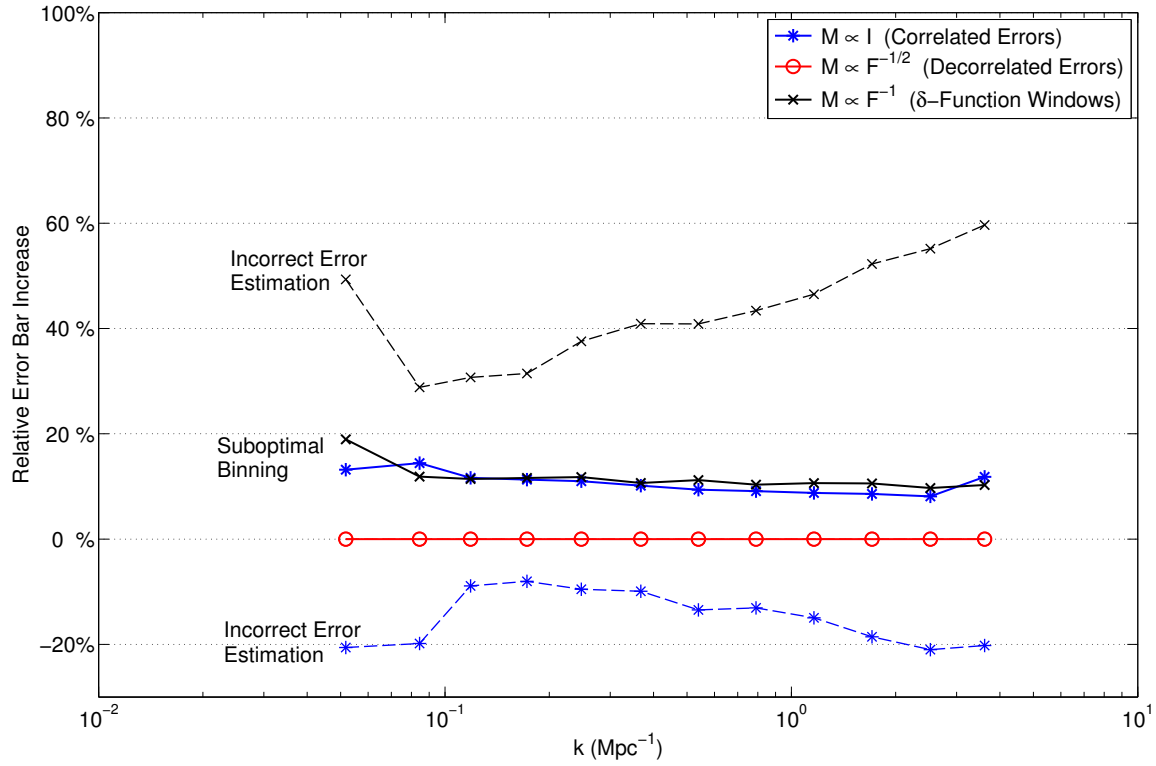
Figure 4-9: Neglecting the fact that the covariance of the power spectrum estimator is, in general, non-diagonal, can lead to two mistakes that can either unnecessarily enlarge our error bars or, even worse, unjustifiably shrink them. In this figure, we first show an approximately 10% increase in the vertical error bars on the power spectrum (solid lines) from a suboptimal inverse variance weighted binning scheme, rather than the inverse covariance weighted binning of Equation (4.34). This problem is obviated by choosing an estimator with decorrelated errors and thus a diagonal covariance matrix. If one simply assumes that the estimator covariance in Equation (4.35) is diagonal when it is not (dotted lines), one is led, depending on the choice of estimator, either to roughly 50% larger error bars than necessary or, worse yet, artificially small error bars. The last mistake, choosing an estimator with small error bars—despite its wide window functions—and then neglecting the off-diagonal terms in the estimator covariance, is potentially the most pernicious since it could lead to a claimed detection in the absence of signal.
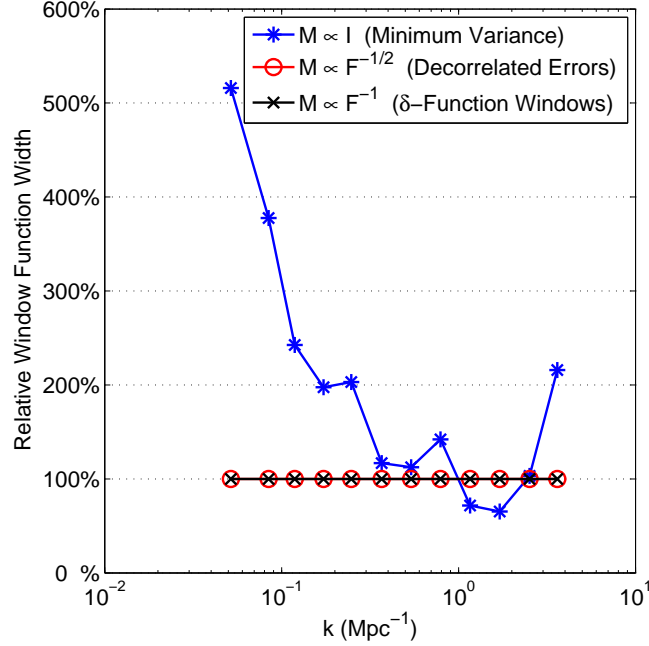
Figure 4-10: Just as with the error bars in Figure 4-9, generating suboptimally binned spherical power spectrum estimates by neglecting off-diagonal terms in the estimator covariance can lead to wider window functions than necessary. We illustrate the effect by comparing the width of the window functions between the 20th and 80th percentiles between the two binning schemes. This is important for the choice of power spectrum estimator with the smallest error bars and widest window functions ($\mathbf{M} \sim \mathbf{I}$). In the case where our power spectrum estimator has uncorrelated errors, there are no off-diagonal terms in the estimator covariance and both binning schemes are identical. In the case of the estimator with δ-function window functions, suboptimal binning does not affect the window functions—though it still affects the vertical errors (see Figure 4-9).

Chapter 5

Empirical Covariance Modeling for 21 cm Power Spectrum Estimation: A Method Demonstration and New Limits from Early Murchison Widefield Array 128-Tile Data

The content of this chapter was submitted to Physical Review D *on March 10, 2015.*

5.1   Introduction

Tomographic mapping of neutral hydrogen using its 21 cm hyperfine transition has the potential to directly probe the density, temperature, and ionization of the intergalactic medium (IGM), from redshift 50 (and possibly earlier) through the end of reionization at $z \sim 6$. This unprecedented view of the so-called "cosmic dawn" can tightly constrain models of the first stars and galaxies [74, 154, 188, 127] and eventually yield an order of magnitude more precise test of the standard cosmological model (ΛCDM) than current probes [135].

Over the past few years, first generation instruments have made considerable

progress toward the detection of the power spectrum of the 21 cm emission during the epoch of reionization (EoR). Telescopes such as the Low Frequency Array (LOFAR [76]), the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER [171]), the Giant Metrewave Radio Telescope (GMRT [167]), and the Murchison Widefield Array (MWA [129, 220, 29]) are now operating, and have begun to set limits on the power spectrum. GMRT set some of the earliest limits [167] and both PAPER [105] and the MWA [59] have presented upper limits across multiple redshifts using small prototype arrays. PAPER has translated its results into a constraint on the heating of the IGM by the first generation of x-ray binaries and miniquasars [173] and has placed the tightest constraints so far on the power spectrum [6] and the thermal history of the IGM [185].

Despite recent advances, considerable analysis challenges remain. Extracting the subtle cosmological signal from the noise is expected to require thousand hour observations across a range of redshifts [151, 25, 117, 87, 170, 218]. Even more daunting is the fact that the 21 cm signal is probably at least 4 orders of magnitude dimmer than the astrophysical foregrounds—due to synchrotron radiation from both our Galaxy and from other galaxies [53, 107, 18, 182, 244, 109].

Recently, simulations and analytical calculations have established the existence of a region in cylindrical Fourier space—in which three-dimensional (3D) Fourier modes $\vec{k}$ are binned into $k_\parallel$ modes along the line of sight and $k_\perp$ modes perpendicular to it—called the "EoR window" that should be fairly free of foreground contamination [50, 172, 230, 156, 89, 225, 218, 125, 126]. Observations of the EoR window confirm that it is largely foreground-free [182, 59] up to the sensitivity limits of current experiments. The boundary of the EoR window is determined by the volume and resolution of the observation, the intrinsic spectral structure of the foregrounds, and the so-called "wedge."

Physically, the wedge arises from the frequency dependence of the point spread function (PSF) of any interferometer, which can create spectral structure from spectrally smooth foregrounds in our 3D maps (see [125] for a rigorous derivation). Fortunately, in $k_\parallel$-$k_\perp$ space, instrumental chromaticity from flat-spectrum sources is re-

stricted to the region below

$$k_\parallel = \theta_0 \frac{D_M(z)E(z)}{D_H(1+z)} k_\perp, \qquad (5.1)$$

where $D_H \equiv c/H_0$, $E(z) \equiv \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$, and $D_m(z) \equiv \int_0^z \mathrm{d}z'/E(z')$ with cosmological parameters from [178]. The size of the region is determined by θ_0, the angle from zenith beyond which foregrounds do not significantly contribute. While most of the foreground emission we observe should appear inside the main lobe of the primary beam, foreground contamination from sources in the sidelobes are also significant compared to the signal [180, 219, 219]. A conservative choice of θ_0 is therefore $\pi/2$, which reflects the fact that the maximum possible delay a baseline can measure corresponds to a source at the horizon [172]. Still, this foreground isolation is not foolproof and can be easily corrupted by miscalibration and imperfect data reduction. Further, slowly varying spectral modes just outside the wedge are also affected when the foreground residuals have spectral structure beyond that imprinted by the chromaticity of the interferometer.

To confidently detect the 21 cm EoR power spectrum, we need rigorous statistical techniques that incorporate models of the cosmological signal, the foregrounds, the instrument, the instrumental noise, and the exact mapmaking procedure. With this information, one may use estimators that preserve as much cosmological information as possible and thoroughly propagate errors due to noise and foregrounds through the analysis pipeline.

The development of such statistical techniques has progressed rapidly over the past few years. The quadratic estimator formalism was adapted [120] from previous work on the cosmic microwave background [208] and galaxy surveys [213]. It was accelerated to meet the data volume challenges of 21 cm tomography [58] and refined to overcome some of the difficulties of working with real data [59]. Further, recent work has shown how to rigorously incorporate the interferometric effects that create the wedge [125, 126, 61], though they rely on precision instrument modeling, including exact per-frequency and per-antenna primary beams and complex gains. A similar

technique designed for drift-scanning telescopes using spherical harmonic modes was developed in [197, 198], which also demonstrated the need for a precise understanding of one's instrument.

However, at this early stage in the development of 21 cm cosmology, precision instrument characterization remains an active area of research [204, 157, 158, 181]. We thus pursue a more cautious approach to foreground modeling that reflects our incomplete knowledge of the instrument by modeling the residual foreground covariance from the data itself. As we will show, this mitigates systematics such as calibration errors that would otherwise impart spectral structure onto the foregrounds, corrupting the EoR window. While not a fully Bayesian approach like those of [206] and [205], our technique discovers both the statistics of the foregrounds and the power spectrum from the data. Our foreground models are subject to certain prior assumptions but are allowed to be data motivated in a restricted space. However, by working in the context of the quadratic estimator formalism, we can benefit from the computational speedups of [58]. This work is meant to build on those techniques and make them more easily applied to real and imperfect data.

This paper is organized into two main parts. In Section 5.2 we discuss the problem of covariance modeling in the context of power spectrum estimation and present a method for the empirical estimation of that foreground model, using MWA data to illustrate the procedure. Then, in Section 8.3, we explain how these data were taken and reduced into maps and present the results of our power spectrum estimation procedure on a few hours of MWA observation, including limits on the 21 cm power spectrum.

5.2 Empirical Covariance Modeling

Before presenting our method of empirically modeling the statistics of residual foregrounds in our maps, we need to review the importance of these covariances to power spectrum estimation. We begin in Section 5.2.1 with a brief review of the quadratic estimator formalism for optimal power spectrum estimation and rigorous error quan-

tification. We then discuss in Section 5.2.2 the problem of covariance modeling in greater detail, highlighting exactly which unknowns we are trying to model with the data. Next we present in Section 5.2.3 our empirical method of estimating the covariance of foreground residuals, illustrated with an application to MWA data. Lastly, we review in Section 5.2.4 the assumptions and caveats that we make or inherit from previous power spectrum estimation work.

5.2.1 Quadratic Power Spectrum Estimator Review

The fundamental goal of power spectrum estimation is to reduce the volume of data by exploiting statistical symmetries while retaining as much information as possible about the cosmological power spectrum [208]. We seek to estimate a set of band powers $\mathbf{p}$ using the approximation that

$$P(\mathbf{k}) \approx \sum_\alpha p_\alpha \chi_\alpha(\mathbf{k}), \tag{5.2}$$

where $P(\mathbf{k})$ is the power spectrum as a function of wave vector $\mathbf{k}$ and χ_α is an indicator function that equals 1 wherever we are approximating $P(\mathbf{k})$ by p_α and vanishes elsewhere.

Following [120, 58, 59], we estimate power spectra from a "data cube"—a set of sky maps of brightness temperature at many closely spaced frequencies—which we represent as a single vector $\hat{\mathbf{x}}$ whose index iterates over both position and frequency. From $\hat{\mathbf{x}}$, we estimate each band power as

$$\widehat{p}_\alpha = \frac{1}{2} M_{\alpha\beta} \left(\hat{\mathbf{x}}_1 - \boldsymbol{\mu} \right)^\mathsf{T} \mathbf{C}^{-1} \mathbf{C}_{,\beta} \, \mathbf{C}^{-1} \left(\hat{\mathbf{x}}_2 - \boldsymbol{\mu} \right) - b_\alpha. \tag{5.3}$$

Here $\boldsymbol{\mu} = \langle \hat{\mathbf{x}} \rangle$, the ensemble average of our map over many different realizations of the observation, and $\mathbf{C}$ is the covariance of our map,

$$\mathbf{C} = \langle \hat{\mathbf{x}} \hat{\mathbf{x}}^\mathsf{T} \rangle - \langle \hat{\mathbf{x}} \rangle \langle \hat{\mathbf{x}} \rangle^\mathsf{T}. \tag{5.4}$$

$\mathbf{C}_{,\beta}$ is a matrix that encodes the response of the covariance to changes in the true, underlying band powers; roughly speaking, it performs the Fourier transforming, squaring, and binning steps one normally associates with computing power spectra.[1] Additionally, $\mathbf{M}$ is an invertible normalization matrix and b_α is the power spectrum bias from nonsignal contaminants in $\hat{\mathbf{x}}$. In this work, we follow [59] and choose a form of $\mathbf{M}$ such that $\mathbf{\Sigma} \equiv \mathrm{Cov}(\hat{\mathbf{p}})$ is diagonal, decorrelating errors in the power spectrum and thus reducing foreground leakage into the EoR window. In order to calculate $\mathbf{M}$ and $\mathbf{\Sigma}$ quickly, we use the fast method of [58] which uses fast Fourier transforms and Monte Carlo simulations to approximate these matrices.

Finally, temporally interleaving the input data into two cubes $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ with the same sky signal but independent noise avoids a noise contribution to the bias b_α as in [59]. Again following [59], we abstain from subtracting a foreground residual bias in order to avoid any signal loss (as discussed in 5.2.3.3).

5.2.2 What Does Our Covariance Model Represent?

Our brightness temperature data cubes are made up of contributions from three statistically independent sources: the cosmological signal, $\hat{\mathbf{x}}^S$; the astrophysical foregrounds, $\hat{\mathbf{x}}^{FG}$; and the instrumental noise $\hat{\mathbf{x}}^N$. It follows that the covariance matrix is equal to the sum of their separate covariances:

$$\mathbf{C} = \mathbf{C}^S + \mathbf{C}^{FG} + \mathbf{C}^N. \tag{5.5}$$

Hidden in the statistical description of the different contributions to our measurement is an important subtlety. Each of these components is taken to be a particular instantiation of a random process, described by a mean and covariance. In the case of the cosmological signal, it is the underlying statistics—the mean and covariance— which encode information about the cosmology and astrophysics. However, we can only learn about those statistics by assuming statistical isotropy and homogeneity and by assuming that spatial averages can stand in for ensemble averages in large

---

[1]For a derivation of an explicit form of $\mathbf{C}_{,\beta}$, see [120] or [58].

volumes. In the case of the instrumental noise, we usually think of the particular instantiation of the noise that we see as the result of a random trial.

The foregrounds are different. There is only one set of foregrounds, and they are not random. If we knew exactly how the foregrounds appear in our observations, we would subtract them from our maps and then ignore them in this analysis. We know that we do not know the foregrounds exactly, and so we choose to model them with our best guess, $\boldsymbol{\mu}^{FG}$. If we define the cosmological signal to consist only of fluctuations from the brightness temperature of the global 21 cm signal, then the signal and the noise both have $\boldsymbol{\mu}^S = \boldsymbol{\mu}^N = 0$. Therefore, we start our power spectrum estimation using Equation (5.3) by subtracting off our best guess as to the foreground contamination in our map. But how wrong are we?

The short answer is that we do not really know that either. But, if we want to take advantage of the quadratic estimator formalism to give the highest weight to the modes we are most confident in, then we must model the statistics of our foreground residuals. If we assume that our error is drawn from some correlated Gaussian distribution, then we should use that *foreground uncertainty covariance* as the proper $\mathbf{C}^{FG}$ in Equation (5.3).

So what do we know about the residual foregrounds in our maps? In theory, our dirty maps are related to the true sky by a set of point spread functions that depend on both position and frequency [61]. This is the result of both the way our interferometer turns the sky into measured visibilities and the way we make maps to turn those visibilities into $\hat{\mathbf{x}}$. In other words, there exists some matrix of PSFs, $\mathbf{P}$ such that

$$\langle \hat{\mathbf{x}} \rangle = \mathbf{P}\mathbf{x}^{\text{true}}. \tag{5.6}$$

The spectral structure in our maps that creates the wedge feature in the power spectrum is a result of $\mathbf{P}$.

We can describe our uncertainty about the true sky—about the positions, fluxes, and spectral indices of both diffuse foregrounds and points sources—with a covariance

matrix $\mathbf{C}^{FG,\text{true}}$ [120, 58], so that

$$\mathbf{C}^{FG} = \mathbf{P}\mathbf{C}^{FG,\text{true}}\mathbf{P}^{\mathsf{T}}. \qquad (5.7)$$

This equation presents us with two ways of modeling the foregrounds. If we feel that we know the relationship between our dirty maps and the true sky precisely, then we can propagate our uncertainty about a relatively small number of foreground parameters, as discussed by [120] and [58], through the $\mathbf{P}$ matrix to get $\mathbf{C}^{FG}$. This technique, suggested by [61], relies on precise knowledge of $\mathbf{P}$. Of course, the relationship between the true sky and our visibility data depends both on the design of our instrument and on its calibration. If our calibration is very good—if we really understand our antenna gains and phases, our primary beams, and our bandpasses—then we can accurately model $\mathbf{P}$.

If we are worried about systematics (and at this early stage of 21 cm tomography with low frequency radio interferometers, we certainly are), then we need a complementary approach to modeling $\mathbf{C}^{FG}$ directly, one that we can use both for power spectrum estimation and for comparison to the results of a more theoretically motivated technique. This is the main goal of this work.

5.2.3 Empirical Covariance Modeling Technique

The idea of using empirically motivated covariance matrices in the quadratic estimator formalism has some history in the field. Previous MWA power spectrum analysis [59] used the difference between time-interleaved data cubes to estimate the overall level of noise, empirically calibrating T_{sys}, the system temperature of the elements. PAPER's power spectrum analysis relies on using observed covariances to suppress systematic errors [173] and on boot-strapped error bars [173, 105]. A similar technique was developed contemporaneously with this work and was used by [6] to estimate covariances.

$\mathbf{C}^{FG}$ has far more elements than we have measured voxels—our cubes have about 2×10^5 voxels, meaning that $\mathbf{C}^{FG}$ has up to 2×10^{10} unique elements. Therefore,

any estimate of $\mathbf{C}^{FG}$ from the data needs to make some assumptions about the structure of the covariance. Since foregrounds have intrinsically smooth spectra, and since one generally attempts to model and subtract smooth spectrum foregrounds, it follows that foreground residuals will be highly correlated along the line of sight. After all, if we are undersubtracting foregrounds at one frequency, we are probably undersubtracting at nearby frequencies too. We therefore choose to focus on empirically constructing the part of $\mathbf{C}^{FG}$ that corresponds to the frequency-frequency covariance—the covariance along the line of sight. If there are n_f frequency channels, then that covariance matrix is only $n_f \times n_f$ elements and is likely dominated by a relatively small number of modes.

In this section, we will present an approach to solving this problem in a way that faithfully reflects the complex spectral structure introduced by an (imperfectly calibrated) interferometer on the bright astrophysical foregrounds. As a worked example, we use data from a short observation with the MWA which we will describe in detail in Section 8.3. We begin with a uniformly weighted map of the sky at each frequency, a model for both point sources and diffuse emission imaged from simulated visibilities, and a model for the noise in each uv cell as a function of frequency.

The idea to model $\mathbf{C}^{FG}$ empirically was put forward by Liu [119]. He attempted to model each line of sight as statistically independent and made no effort to separate $\mathbf{C}^{FG}$ from $\mathbf{C}^{N}$ or to reduce the residual noisiness of the frequency-frequency covariance.

Our approach centers on the idea that the covariance matrix can be approximated as block diagonal in the uv basis of Fourier modes perpendicular to the line of sight. In other words, we are looking to express $\mathbf{C}^{FG}$ as

$$C^{FG}_{uu'vv'ff'} \approx \delta_{uu'}\delta_{vv'}\widehat{C}_{ff'}(k_\perp), \qquad (5.8)$$

where $k_\perp$ is a function of $\sqrt{u^2 + v^2}$. This is the tensor product of our best guess of the frequency-frequency covariance $\widehat{\mathbf{C}}$ and the identity in both Fourier coordinates perpendicular to the line of sight. In this way, we can model different frequency-

frequency covariances as a function of $|u|$ or equivalently, $k_\perp$, reflecting that fact that the wedge results from greater leakage of power up from low $k_\parallel$ as one goes to higher $k_\perp$. This method also has the advantage that $\mathbf{C}$ becomes efficient to both write down and invert directly, removing the need for the preconditioned conjugate gradient algorithm employed by [58].

This approximation is equivalent to the assumption that the residuals in every line of sight are statistically independent of position. This is generally a pretty accurate assumption as long as the primary beam does not change very much over the map from which we estimate the power spectrum. However, because $\widehat{C}_{ff'}(k_\perp)$ depends on the angular scale, we are still modeling correlations that depend only on the distance between points in the map.

While we might expect that the largest residual voxels correspond to errors in subtracting the brightest sources, the voxels in the residual data cube (the map minus the model) are only weakly correlated with the best-guess model of the foregrounds (we find a correlation coefficient $\rho = 0.116$, which suggests that sources are removed to roughly the 10% level, assuming that undersubtraction dominates). As we improve our best guess of the model foregrounds through better deconvolution, we expect ρ to go down, improving the assumption that foregrounds are block diagonal in the uv basis. We will now present the technique we have devised in four steps, employing MWA data as a method demonstration.

5.2.3.1 Compute sample covariances in uv annuli.

We begin our empirical covariance calculation by taking the residual data cubes, defined as

$$\hat{\mathbf{x}}^{\mathrm{res}} \equiv \hat{\mathbf{x}}_1/2 + \hat{\mathbf{x}}_2/2 - \boldsymbol{\mu}, \tag{5.9}$$

and performing a discrete Fourier transform[2] at each frequency independently to get $\widetilde{\mathbf{x}}^{\mathrm{res}}$. This yields $n_x \times n_y$ sample "lines of sight" (uv cells for all frequencies), as

---

[2] For simplicity, we used the unitary discrete Fourier transform for these calculations and ignore any factors of length or inverse length that might come into these calculations only to be canceled out at a later step.

many as we have pixels in the map. As a first step toward estimating $\widehat{\mathbf{C}}$, we use the unbiased sample covariance estimator from these residual lines of sight. However, instead of calculating a single frequency-frequency covariance, we want to calculate many different $\widehat{\mathbf{C}}^{\mathrm{res}}$ matrices to reflect the evolution of spectral structure with $k_{\perp}$ along the wedge. We therefore break the uv plane into concentric annuli of equal width and calculate $\widehat{\mathbf{C}}^{\mathrm{res}}_{uv}$ for each uv cell as the sample covariance of the $N^{\mathrm{LOS}} - 2$ lines of sight in that annulus, excluding the cell considered and its complex conjugate. Since the covariance is assumed to be block diagonal, this eliminates a potential bias that comes from downweighting a uv cell using information about that cell. Thus,

$$\widehat{C}^{\mathrm{res}}_{uv,ff'} = \sum_{\substack{\text{other } u',v' \\ \text{in annulus}}} \frac{\left(\widetilde{x}^{\mathrm{res}}_{u'v'f} - \langle \widetilde{x}^{\mathrm{res}}_f \rangle\right)\left(\widetilde{x}^{\mathrm{res}}_{u'v'f'} - \langle \widetilde{x}^{\mathrm{res}}_{f'} \rangle\right)^*}{N^{\mathrm{LOS}} - 2 - 1}, \tag{5.10}$$

where $\langle \widetilde{x}^{\mathrm{res}}_f \rangle$ is an average over all u' and v' in the annulus. We expect this procedure to be particularly effective in our case because the uv coverage of the MWA after rotation synthesis is relatively symmetric.

As a sense check on these covariances, we plot their largest 30 eigenvalues in Figure 5-1. We see that as $|u|$ (and thus $k_{\perp}$) increases, the eigenspectra become shallower. At high $k_{\perp}$, the effect of the wedge is to leak power to a range of $k_{\parallel}$ values. The eigenspectrum of intrinsically smooth foregrounds should be declining exponentially [121]. The wedge softens that decline. These trends are in line with our expectations and further motivate our strategy of forming covariance matrices for each annulus independently.

Because we seek only to estimate the foreground portion of the covariance, the formal rank deficiency of $\widehat{\mathbf{C}}^{\mathrm{res}}_{uv}$ is not a problem.[3] All we require is that the largest (and thus more foreground-dominated) modes be well measured. In this analysis, we used six concentric annuli to create six different frequency-frequency foreground covariances. Using more annuli allows for better modeling of the evolution of the wedge with $k_{\perp}$ at the expense of each estimate being more susceptible to noise and rank deficiency.

---

[3]In fact, the rank of $\widehat{\mathbf{C}}^{\mathrm{res}}_{uv}$ is $N_{\mathrm{LOS}} - 3$ if $N^{\mathrm{LOS}} - 2 \leq n_f$.
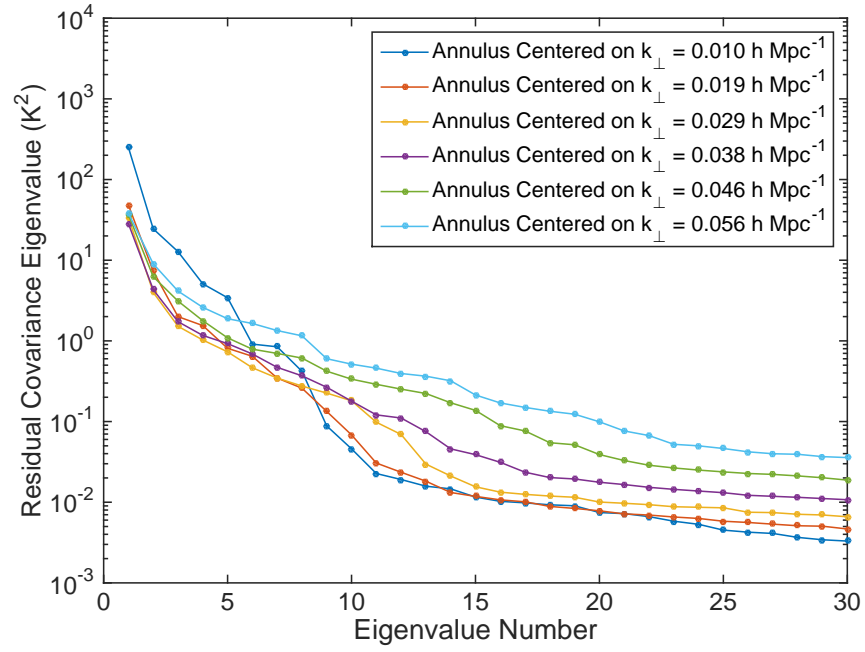
Figure 5-1: The evolution of the wedge with $k_\perp$ motivates us to model foregrounds separately for discrete values of $k_\perp$. In this plot of the 30 largest eigenvalues of the observed residual covariance (which should include both noise and foregrounds) sampled in six concentric annuli, we see steeper declines toward a noise floor for the inner annuli than the outer annuli. This is consistent with the expected effect of the wedge—higher $k_\perp$ modes should be foreground contaminated at higher $k_\parallel$.

5.2.3.2 Subtract the properly projected noise covariance.

The covariances computed from these uv lines of sight include contributions from the 21 cm signal and instrumental noise as well as foregrounds. We can safely ignore the signal covariance for now as we are far from the regime where sample variance is significant. We already have a theoretically motivated model for the noise (based on the uv sampling) that has been empirically rescaled to the observed noise in the difference of time-interleaved data (the same basic procedure as in [59]). We would like an empirical estimate of the residual foreground covariance alone to use in $\mathbf{C}^{FG}$ and thus must subtract off the part of our measurement we think is due to noise variance.

To get to $\widehat{\mathbf{C}}_{uv}^{\mathrm{FG}}$ from $\widehat{\mathbf{C}}_{uv}^{\mathrm{res}}$, we subtract our best guess of the portion of $\widehat{\mathbf{C}}_{uv}^{\mathrm{res}}$ that is due to noise, which we approximate by averaging the noise model variances in all the other uv cells in the annulus at that given frequency, yielding

$$\widehat{C}_{uv,ff'}^{\mathrm{N}} = \frac{1}{N^{\mathrm{LOS}}} \sum_{\substack{\text{other } u',v' \\ \text{in annulus}}} \delta_{uu'}\delta_{vv'}\delta_{ff'}C_{uu'vv'ff'}^{N}. \tag{5.11}$$

Note, however, that $\widehat{\mathbf{C}}_{uv}^{\mathrm{N}}$ is full rank while $\widehat{\mathbf{C}}_{uv}^{res}$ is typically rank deficient. Thus a naive subtraction would oversubtract the noise variance in the part of the subspace of $\widehat{\mathbf{C}}_{uv}^{\mathrm{N}}$ where $\widehat{\mathbf{C}}_{uv}^{res}$ is identically zero. Instead, the proper procedure is to find the projection matrices $\mathbf{\Pi}_{uv}$ that discard all eigenmodes outside the subspace where $\widehat{\mathbf{C}}_{uv}^{\mathrm{res}}$ is full rank. Each should have eigenvalues equal to zero or one only and have the property that

$$\mathbf{\Pi}_{uv}\widehat{\mathbf{C}}_{uv}^{\mathrm{res}}\mathbf{\Pi}_{uv}^{\mathsf{T}} = \widehat{\mathbf{C}}_{uv}^{\mathrm{res}}. \tag{5.12}$$

Only after projecting out the part of $\widehat{\mathbf{C}}_{uv}^{N}$ inside the unsampled subspace can we self-consistently subtract our best guess of the noise contribution to the subspace in which we seek to estimate foregrounds. In other words, we estimate $\widehat{\mathbf{C}}_{uv}^{\mathrm{FG}}$ as

$$\widehat{\mathbf{C}}_{uv}^{\mathrm{FG}} = \widehat{\mathbf{C}}_{uv}^{\mathrm{res}} - \mathbf{\Pi}_{uv}\widehat{\mathbf{C}}_{uv}^{\mathrm{N}}\mathbf{\Pi}_{uv}^{\mathsf{T}}. \tag{5.13}$$
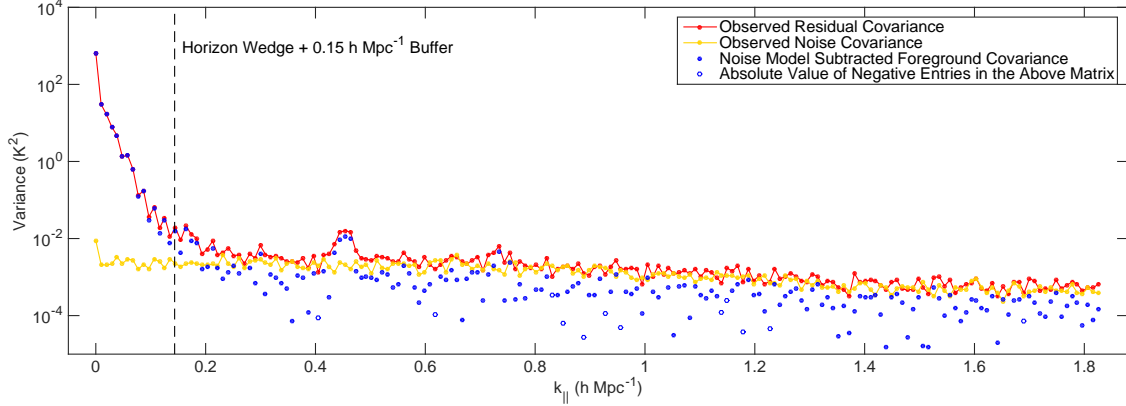
Figure 5-2: Examining the diagonal elements of the observed residual and inferred foreground covariance matrices in Fourier space reveals the effectiveness of subtracting model for the noise covariance. In red, we plot the observed residual covariance, which contains both foregrounds and noise. As a function of $k_\parallel$, the two separate relatively cleanly—there is a steeply declining foreground portion on the left followed by a relatively flat noise floor on the right. The theory that the right-hand portion is dominated by noise is borne out by the fact that it so closely matches the observed noise covariance, inferred lines of sight of $\mathbf{x}_1 - \mathbf{x}_2$, which should have only noise and no sky signal at all. The regions where they differ significantly, for example at $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$, are attributable to systematic effects like the MWA's coarse band structure that have not been perfectly calibrated out. For the example covariances shown here (which correspond to a mode in the annulus at $k_\perp \approx 0.010\,h\,\mathrm{Mpc}^{-1}$), we can see that subtracting a properly projected noise covariance removes most of the power from the noise-dominated region, leaving only residual noise that appears both as negative power (open blue circle) and as positive power (closed blue circles) at considerably lower magnitude.

We demonstrate the effectiveness of this technique in Figure 5-2 by plotting the diagonal elements of the Fourier transform of $\widehat{\mathbf{C}}_{uv}^{\mathrm{res}}$ and $\widehat{\mathbf{C}}_{uv}^{\mathrm{FG}}$ along the line of sight. Subtracting of the noise covariance indeed eliminates the majority of the power in the noise dominated modes at high $k_\parallel$; thus we expect it also to fare well in the transition region near the edge of the wedge where foreground and noise contributions are comparable.

5.2.3.3   Perform a $k_\parallel$ filter on the covariance.

Despite the relatively clean separation of foreground and noise eigenvalues, inspection of some of the foreground-dominated modes in the top panel of Figure 5-3 reveals

residual noise. Using a foreground covariance constructed from these noisy foreground eigenmodes to downweight the data during power spectrum estimation would errantly downweight some high $k_\parallel$ modes in addition to the low $k_\parallel$ foreground-dominated modes. To avoid this double counting of the noise, we allow the foreground covariance to include only certain $k_\parallel$ modes by filtering $\widehat{\mathbf{C}}^{\mathrm{FG}}_{uv}$ in Fourier space to get $\widehat{\mathbf{C}}^{\mathrm{FG,filtered}}_{uv}$. Put another way, we are imposing a prior on which Fourier modes we think have foreground power in them. The resulting noise filtered eigenmodes are shown in the bottom panel of Figure 5-3.

In practice, implementing this filter is subtle. We interpolate $\widehat{\mathbf{C}}^{\mathrm{FG}}$ over the flagged frequency channels using a cubic spline, then symmetrically pad the covariance matrix, forcing its boundary condition to be periodic. We then Fourier transform, filter, inverse Fourier transform, remove the padding, and then rezero the flagged channels.

Selecting a filter to use is also a subtle choice. We first keep modes inside the horizon wedge with an added buffer. For each annulus, we calculate a mean value of $k_\perp$, and then use Equation (7.21) to calculate the $k_\parallel$ value of the horizon wedge, using $\theta_0 = \pi/2$. Although the literature suggests a 0.1 to $0.15\,h\,\mathrm{Mpc}^{-1}$ buffer for "suprahorizon emission" due to some combination of intrinsic spectral structure of foregrounds, primary beam chromaticity, and finite bandwidth [182, 184], we pick a conservative $0.5\,h\,\mathrm{Mpc}^{-1}$. Then we examine the diagonal of $\widehat{\mathbf{C}}^{\mathrm{FG}}$ (Figure 5-2) to identify additional foreground modes, this time in the EoR window, due to imperfect bandpass calibration appearing as spikes. One example is the peak at $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$. Such modes contribute errant power to the EoR window at constant $k_\parallel$. Since these modes result from the convolution of the foregrounds with our instrument, they also should be modeled in $\mathbf{C}^{FG}$ in order to minimize their leakage into the rest of the EoR window.

One might be concerned that cosmological signal and foregrounds theoretically both appear in the estimate of $\mathbf{C}^{FG}$ that we have constructed, especially with our conservative $0.5\,h\,\mathrm{Mpc}^{-1}$ buffer that allows foregrounds to be discovered well into the EoR window. For the purposes of calculating $\mathbf{C}^{-1}(\widehat{\mathbf{x}} - \boldsymbol{\mu})$ in the quadratic estimator in Equation (5.3), that is fine since its effect is to partially relax the assumption that sample variance can be ignored. However, the calculation of the bias depends on
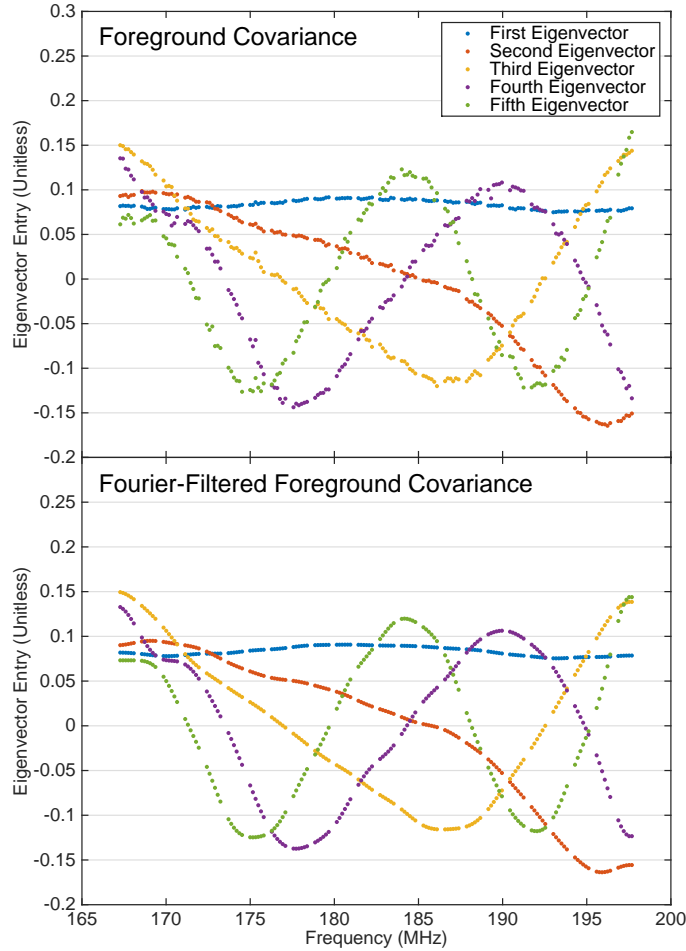
Figure 5-3: The foreground covariance we estimate from our limited data set is still very noisy, and we run the risk of overfitting the noise in our measurements if we take it at face value. In the top panel, we plot the eigenvectors corresponding to the five largest eigenvalues of $\widehat{\mathbf{C}}^{\mathrm{FG}}$ for a mode in the annulus centered on $k_\perp \approx 0.010\,h\,\mathrm{Mpc}^{-1}$. In the bottom panel, we show dominant eigenvectors of the Fourier-filtered covariance. As expected, they resemble the first five Fourier modes. The missing data every 1.28 MHz are due to channels flagged at the edge of the coarse bandpass of the MWA's polyphase filter bank—the most difficult part of the band to calibrate.

being able to differentiate signal from contaminants [208, 120, 58].

The noise contribution to the bias can be eliminated by cross-correlating maps made from interleaved time steps [59]. However, we cannot use our inferred $\mathbf{C}^{FG}$ to subtract a foreground bias without signal loss. That said, we can still set an upper limit on the 21 cm signal. By following the data and allowing the foreground covariance to have power inside the EoR window, we are minimizing the leakage of foregrounds into uncontaminated regions and we are accurately marking those regions as having high variance. As calibration and the control of systematic effects improves, we should be able to isolate foregrounds to outside the EoR window, impose a more aggressive Fourier filter on $\mathbf{C}^{FG}$, and make a detection of the 21 cm signal by employing foreground avoidance.

5.2.3.4   Cut out modes attributable to noise.

After suppressing the noisiest modes with our Fourier filter, we must select a cutoff beyond which the foreground modes are irrecoverably buried under noise. We do this by inspecting the eigenspectrum of $\widehat{\mathbf{C}}^{\mathrm{FG,filtered}}_{uv}$. The true $\mathbf{C}^{FG}$, by definition, admits only positive eigenvalues (though some of them should be vanishingly small).

By limiting the number of eigenvalues and eigenvectors we ultimately associate with foregrounds, we also limit the potential for signal loss by allowing a large portion of the free parameters to get absorbed into the contaminant model [207, 6]. When measuring the power spectrum inside the EoR window, we can be confident that signal loss is minimal compared to foreground bias and other errors.

We plot in Figure 5-4 the eigenspectra of $\widehat{\mathbf{C}}^{\mathrm{res}}_{uv}$, $\widehat{\mathbf{C}}^{\mathrm{FG}}_{uv}$, and $\widehat{\mathbf{C}}^{\mathrm{FG,filtered}}_{uv}$, sorted by absolute value. There are two distinct regions—the sharply declining foreground-dominated region and a flatter region with many negative eigenvalues. We excise eigenvectors whose eigenvalues are smaller in absolute value than the most negative eigenvalue. This incurs a slight risk of retaining a few noise dominated modes, albeit strongly suppressed by our noise variance subtraction and our Fourier filtering. Finally we are able to construct the full covariance $\widehat{\mathbf{C}}$ using Equation (5.8).
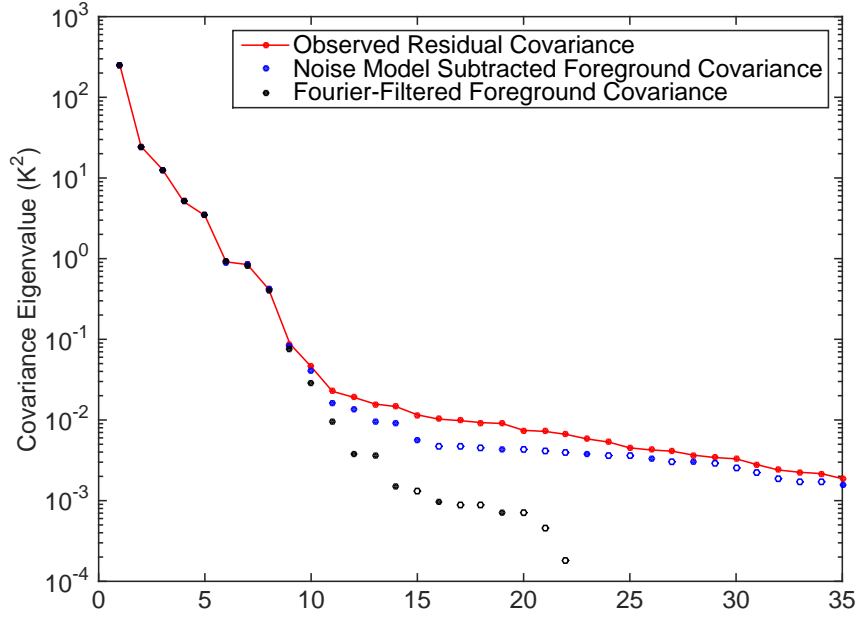
Figure 5-4: The evolution of the eigenvalues of our estimated foreground covariance matrix for a mode in the annulus corresponding to $k_\perp \approx 0.010\,h\,\mathrm{Mpc}^{-1}$ at each of the first three stages of covariance estimation. First we calculate a sample covariance matrix from the residual data cubes (shown in red). Next we subtract our best guess as to the part of the diagonal of that matrix that originates from instrumental noise, leaving the blue dots (open circles are absolute values of negative eigenvalues). Then we filter out modes in Fourier space along the line of sight that we think should be noise dominated, leaving the black dots. Finally, we project out the eigenvectors associated with eigenvalues whose magnitude is smaller than the largest negative eigenvalue, since those are likely due to residual noise. What remains is our best guess at the foreground covariance in an annulus and incorporates as well as possible our prior beliefs about its structure.

5.2.4 Review of Assumptions and Caveats

Before proceeding to demonstrate the effectiveness of our empirical covariance modeling method, it is useful to review and summarize the assumptions made about map-making and covariance modeling. Some are inherited from the previous application of quadratic power spectrum estimation to the MWA [59], while others are necessitated by our new, more faithful foreground covariance. Relaxing these assumptions in a computationally efficient manner remains a challenge we leave for future work.

i. We adopt the flat sky approximation as in [58, 59], allowing us to use the fast Fourier transform to quickly compute power spectra. The error incurred from this approximation on the power spectrum is expected to be smaller than 1% [59].

ii. We assume the expectation value of our uniformly weighted map is the true sky (i.e., $\langle \widehat{\mathbf{x}} \rangle = \mathbf{x}^{\text{true}}$) when calculating $\mathbf{C}_{,\beta}$ in Equation 5.3, again following [59]. In general $\langle \widehat{\mathbf{x}} \rangle$ is related to $\mathbf{x}^{\text{true}}$ by $\mathbf{P}$, the matrix of point spread functions [61]. Here we effectively approximate the PSF as position independent. Relaxing this approximation necessitates the full mapmaking theory presented in [61] which has yet to be integrated into a power spectrum estimation pipeline.

iii. We approximate the foreground covariance as uncorrelated between different uv cells (and thus block diagonal). At some level there likely are correlations in uv, though those along the line of sight are far stronger. It may be possible to attempt to calculate these correlations empirically, but it would be very difficult considering relative strength of line-of-sight correlations. It may also be possible to use a nonempirical model, though that has the potential to make the computational speedups of [58] more difficult to attain.

iv. We approximate the frequency-frequency foreground covariance as constant within each annulus, estimating our covariance for each uv cell only from other cells in the same annulus. In principle, even if the foreground residuals were isotropic,

there should be radial evolution within each annulus which we ignore for this analysis.

v. The Fourier filter is a nontrivial data analysis choice balancing risk of noise double counting against that of insufficiently aggressive foreground downweighting.

vi. In order to detect the 21 cm signal, we assume that foregrounds can be avoided by working within the EoR window. Out of fear of losing signal, we make no effort to subtract a residual foreground bias from the window. This makes a detection inside the wedge impossible and it risks confusing foreground contamination in the window for a signal. Only analysis of the dependence of the measurement on z, k, $k_{\parallel}$, and $k_{\perp}$ can distinguish between systematics and the true signal.

5.3 Results

We can now demonstrate the statistical techniques we have motivated and developed in Section 5.2 on the problem of estimating power spectra from a 3 h observation with the 128-antenna MWA. We begin with a discussion of the instrument and the observations in Section 5.3.1. In Section 5.3.2 we detail the data processing from raw visibilities to calibrated maps from which we estimate both the foreground residual covariance matrix and the power spectrum. Finally, in Section 8.3 we present our results and discuss lessons learned looking toward a detection of the 21 cm signal.

5.3.1 Observation Summary

The 128-antenna Murchison Widefield Array began deep EoR observations in mid-2013. We describe here the salient features of the array and refer to [220] for a more detailed description. The antennas are laid out over a region of radius 1.5 km in a quasirandom, centrally concentrated distribution which achieves approximately complete uv coverage at each frequency over several hours of rotation synthesis [15]. Each antenna element is a phased array of 16 wideband dipole antennas whose phased sum forms a discretely steerable 25° beams (full width at half maximum) at 150 MHz

with frequency-dependent, percent level sidelobes [157]. We repoint the beam to our field center on a 30 min cadence to correct for earth rotation, effectively acquiring a series of drift scans over this field.

We observe the MWA "EOR0" deep integration field, centered at R.A.(J2000) $= 0^h\,0^m\,0^s$ and decl.(J2000) $= -30°\,0'\,0''$. It features a near-zenith position, a high Galactic latitude, minimal Galactic emission [54], and an absence of bright extended sources. This last property greatly facilitates calibration in comparison to the "EOR2" field—a field dominated by the slightly resolved radio galaxy Hydra A at its center—which was used by [234] and [59]. A nominal 3 h set of EOR0 observations was selected during the first weeks of observing to use for refining and comparing data processing, imaging, and power spectra pipelines [102]. In this work, we use the "high band," near-zenith subset of these observations with 30.72 MHz of bandwidth and center frequency of 182 MHz, recorded on Aug 23, 2013 between 16:47:28 and 19:56:32 UTC (22.712 and 1.872 hours LST).

5.3.2 Calibration and Mapmaking Summary

Preliminary processing, including radio frequency interference (RFI) flagging followed by time and frequency averaging, was performed with the `COTTER` package [163] on raw correlator data. These data were collected at 40 kHz resolution with an integration time of 0.5 s, and averaged to 80 kHz resolution with a 2 s integration time to reduce the data volume. Additionally, 80 kHz at the upper and lower edges of each of 24 coarse channels (each of width 1.28 MHz) of the polyphase filter bank is flagged due to known aliasing.

As in [59], we undertake snapshot-based processing in which each minute-scale integration is calibrated and imaged independently. Snapshots are combined only at the last step in forming a Stokes I image cube, allowing us to properly align and weight them despite different primary beams due to sky rotation and periodic repointing. While sources are forward modeled for calibration and foreground subtraction using the full position dependent PSF (i.e., the synthesized beam), we continue to approximate it as position independent (and equal to that of a point source at the

field center) during application of uniform weighting and computation of the noise covariance.

We use the calibration, foreground modeling, and first stage image products produced by the Fast Holographic Deconvolution[4] (FHD) pipeline as described by [102]. The calibration implemented in the FHD package is an adaptation of the fast algorithm presented by [192] with a baseline cutoff of $b > 50\lambda$. In this data reduction, the point source catalogs discussed below are taken as the sky model for calibration. Solutions are first obtained per antenna and per frequency before being constrained to linear phase slopes and quadratic amplitude functions after correcting for a median antenna-independent amplitude bandpass. The foreground model used for subtraction includes models both of diffuse radio emission [14] and point sources. In detail, the point source catalog is the union of a deep MWA point source survey within 20° of the field center [37], the shallower but wider MWA commissioning point source survey [97], and the Culgoora catalog [200]. Note that calibration and foreground subtraction of off-zenith observations are complicated by Galactic emission picked up by primary beam sidelobes, and are active topics of investigation [180, 219, 217]. During these observations a single antenna was flagged due to known hardware problems, and 1–5 more were flagged for any given snapshot due to poor calibration solutions.

These calibration, foreground modeling, and imaging steps constitute notable improvements over [59]. In that work, the presence of the slightly resolved Hydra A in their EOR2 field likely limited calibration and subtraction fidelity as only a point source sky model was used. In contrast, the EOR0 field analyzed here lacks any such nearby radio sources. Our foreground model contains ~ 2500 point sources within the main lobe and several thousand more in the primary beam sidelobes in addition to the aforementioned diffuse map. A last improvement in the imaging is the more frequent interleaving of time steps for the cross power spectrum, which we performed at the integration scale ($2\,\mathrm{s}$) as opposed to the snapshot scale (a few minutes). This ensures that both $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ have identical sky responses and thus allows us

---

[4]For a theoretical discussion of the algorithm see [203]. The code is available at `https://github.com/miguelfmorales/FHD`.

to accurately estimate the noise in the array from difference cubes. Assuming that the system temperature contains both an instrumental noise temperature and a frequency dependent sky noise temperature that scales as $\nu^{-2.55}$, the observed residual root-mean-square brightness temperature is consistent with T_{sys} ranging from $450\,\text{K}$ at $167\,\text{MHz}$ to $310\,\text{K}$ at $198\,\text{MHz}$, in line with expectations [15].

As discussed in [102] and [36], FHD produces naturally weighted sky, foreground model, "weights," and "variances" cubes, as well as beam-squared cubes. All are saved in image space using the HEALPix format [79] with $N_{\text{side}} = 1024$. Note that these image cubes are crops of full-sky image cubes to a $16° \times 16°$ square field of view, as discussed below. The sky, foreground model, and weights cubes are image space representations of the measured visibilities, model visibilities, and sampling function, respectively, all originally gridded in uv space using the primary beam as the gridding kernel. The variances cube is similar to the weights cube, except the gridding kernel is the $s$quare of the uv space primary beam. It represents the proper quadrature summation of independent noise in different visibilities when they contribute to the same uv cell, and will ultimately become our diagonal noise covariance model. The FHD cubes from all ninety-four $112\,\text{s}$ snapshots are optimally combined in this "holographic" frame in which the true sky is weighted by two factors of the primary beam, as in [59].

We perform a series of steps to convert the image cube output of FHD into uniformly weighted Stokes I cubes accompanied by appropriate uv coverage information for our noise model. We first map these data cubes onto a rectilinear grid, invoking the flat sky approximation. We do this by rotating the (RA,Dec) HEALPix coordinates of the EOR0 field to the north pole $(0°,90°)$, and then projecting and gridding onto the xy plane with $0.2° \times 0.2°$ resolution over a $16° \times 16°$ square field of view. To reduce the data volume while maintaining cosmological sensitivity, we coarse grid to approximately $0.5°$ resolution by Fourier transforming and cropping these cubes in the uv plane at each frequency. We form a uniformly weighted Stokes I cube $I_{\text{uni}}(\vec{\theta})$ by first summing the XX and YY data cubes, resulting in a naturally weighted, holographic stokes I cube $I_{\text{nat},h}(\vec{\theta}) = I_{XX,h}(\vec{\theta}) + I_{YY,h}(\vec{\theta})$. Then we divide out the

holographic weights cube $W_h(\vec{\theta})$ in uv space, which applies uniform weighting and removes one image space factor of the beam, and lastly divide out the second beam factor $B(\vec{\theta})$: $I_{\mathrm{uni}}(\vec{\theta}) = \mathcal{F}^{-1}[\mathcal{F}I_{\mathrm{nat},h}(\vec{\theta})/\mathcal{F}W_h(\vec{\theta})]/B(\vec{\theta})$, where $\mathcal{F}$ represents a Fourier transform and $B(\vec{\theta}) = [B_{XX}^2(\vec{\theta}) + B_{YY}^2(\vec{\theta})]^{1/2}$. Consistent treatment of the variances cube requires uv space division of *two* factors of the weights cube followed by image space division of *two* factors of the beam.

Lastly, we frequency average from 80 kHz to 160 kHz, flagging a single 160 kHz channel the edge of each 1.28 MHz coarse channel due to polyphase filter bank attenuation and aliasing, which make these channels difficult to reliably calibrate. Following [59], we also flag poorly observed uv cells and uv cells whose observation times vary widely between frequencies. In all cases, we formally set the variance in flagged channels and uv cells in $\mathbf{C}^N$ to infinity and use the pseudoinverse to project out flagged modes [59].

5.3.3   Power Spectrum Results

We can now present the results of our method applied to 3 h of MWA-128T data. We first study cylindrically averaged, two-dimensional (2D) power spectra and their statistics, since they are useful for seeing the effects of foregrounds and systematic errors on the power spectrum. We form these power spectra with the full 30.72 MHz instrument bandwidth to achieve maximal $k_\parallel$ resolution.

We begin with the 2D power spectrum itself (Figure 5-5) in which several important features can be observed. First, the wedge and EoR window are clearly distinguishable, with foregrounds suppressed by at least 5 orders of magnitude across most of the EoR window. At high $k_\perp$, the edge of the wedge is set by the horizon while at low $k_\perp$ the cutoff is less clear. There appears to be some level of suprahorizon emission, which was also observed with PAPER in [182] and further explained by [125]. Consistent with Figure 5-1 we see the strongest foreground residual power at low $k_\perp$, meaning that there still remains a very large contribution from diffuse emission from our Galaxy—potentially from sidelobes of the primary beam affecting the shortest baselines [219, 217].
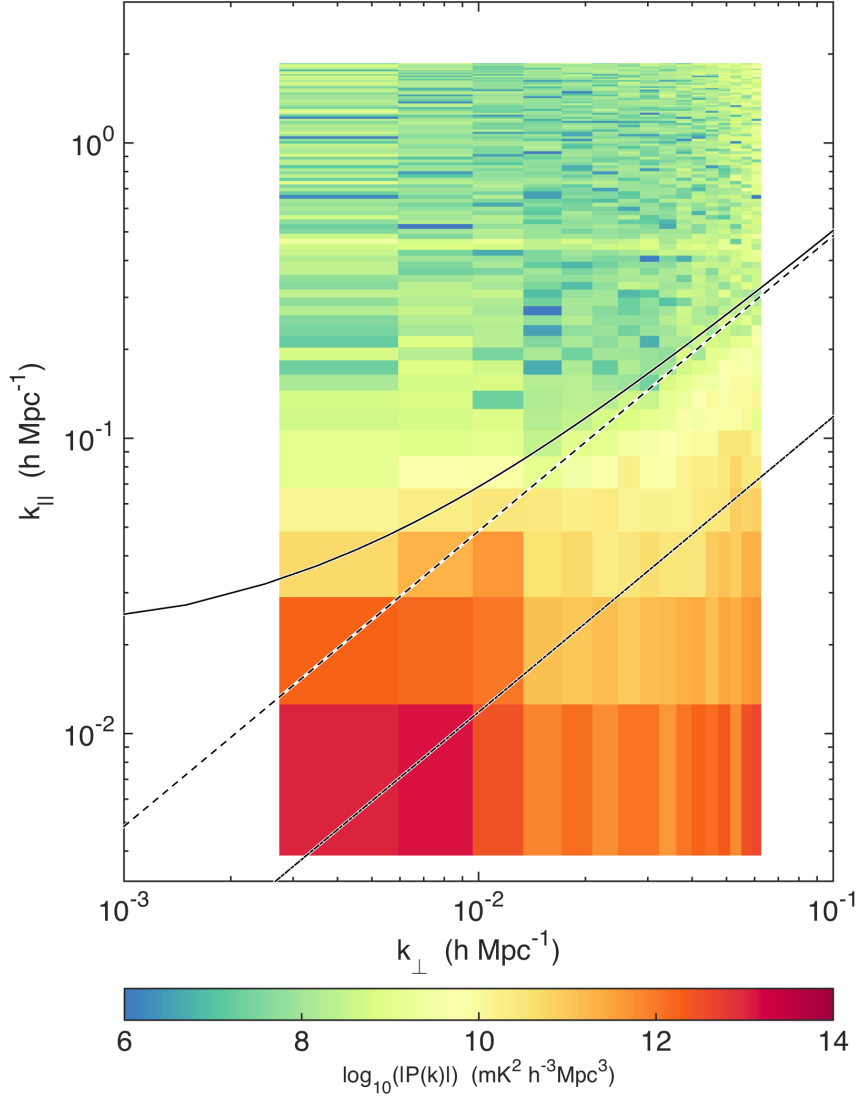
Figure 5-5: Our power spectrum clearly exhibits the typical EoR window structure with orders-of-magnitude suppression of foregrounds in the EoR window. Here we plot our estimates for $|P(k_\perp, k_\parallel)|$ for the full instrumental bandwidth, equivalent to the range $z = 6.2$ to $z = 7.5$. Overplotted is the wedge from Equation 7.21 corresponding to the first null in the primary beam (dash-dotted line), the horizon (dashed line), and the horizon with a relatively aggressive $0.02\,h\,\mathrm{Mpc}^{-1}$ buffer (solid curve). In addition to typical foreground structure, we also see the effect of noise at high and very low $k_\perp$ where baseline coverage is poor. We also clearly see a line of power at constant $k_\parallel \approx 0.45\,h\,\mathrm{Mpc}^{-1}$, attributable to miscalibration of the instrument's bandpass and cable reflections [36].

We also see evidence for less-than-ideal behavior. Through we identified spectral structure appearing at $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$ in Figure 5-2 and included it in our foreground residual covariance, that contamination still appears here as a horizontal line. By including it in the foreground residual model, we increase the variance we associate with those modes and we decrease the leakage out of those modes, isolating the effect to only a few $k_\parallel$ bins.

While Figure 5-5 shows the magnitude of the 2D power spectrum, Figure 5-6 shows its sign using a split color scale, providing another way to assess foreground contamination in the EoR window. Because we are taking the cross power spectrum between two cubes with identical sky signal but independent noise realizations, the noise dominated regions should be positive or negative with equal probability. This is made possible by our use of a power spectrum estimator normalized such that $\boldsymbol{\Sigma} \equiv \mathrm{Cov}(\hat{\mathbf{p}})$ is a diagonal matrix [58]. This choice limits leakage of foreground residuals from the wedge into the EoR window [59].

By this metric, the EoR window is observed to be noise dominated with only two notable exceptions. The first is the region just outside the wedge at low $k_\perp$ attributable to suprahorizon emission due to some combination of intrinsic foreground spectral structure, beam chromaticity, and finite bandwidth. This suggests our aggressive $0.02\,h\,\mathrm{Mpc}^{-1}$ cut beyond the horizon will leave in some foreground contamination when we bin to form one-dimensional (1D) power spectra. As long as we are only claiming an upper limit on the power spectrum, this is fine. A detection of foregrounds is also an upper limit on the cosmological signal. More subtle is the line of positive power at $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$, confirming our hypothesis that the spike observed in Figure 5-5 is indeed an instrumental systematic since it behaves the same way in both time-interleaved data cubes. There is also a hint of a similar effect at $k_\parallel \sim 0.75\,h\,\mathrm{Mpc}^{-1}$, possibly visible in Figure 5-2 as well. We attribute both to bandpass miscalibration due to cable reflections, complicated at these frequency scales by the imperfect channelization of the MWA's two-stage polyphase filter, as well as slight antenna dependence of the bandpass due to cable length variation [36].

Additionally, the quadratic estimator formalism relates our covariance models of
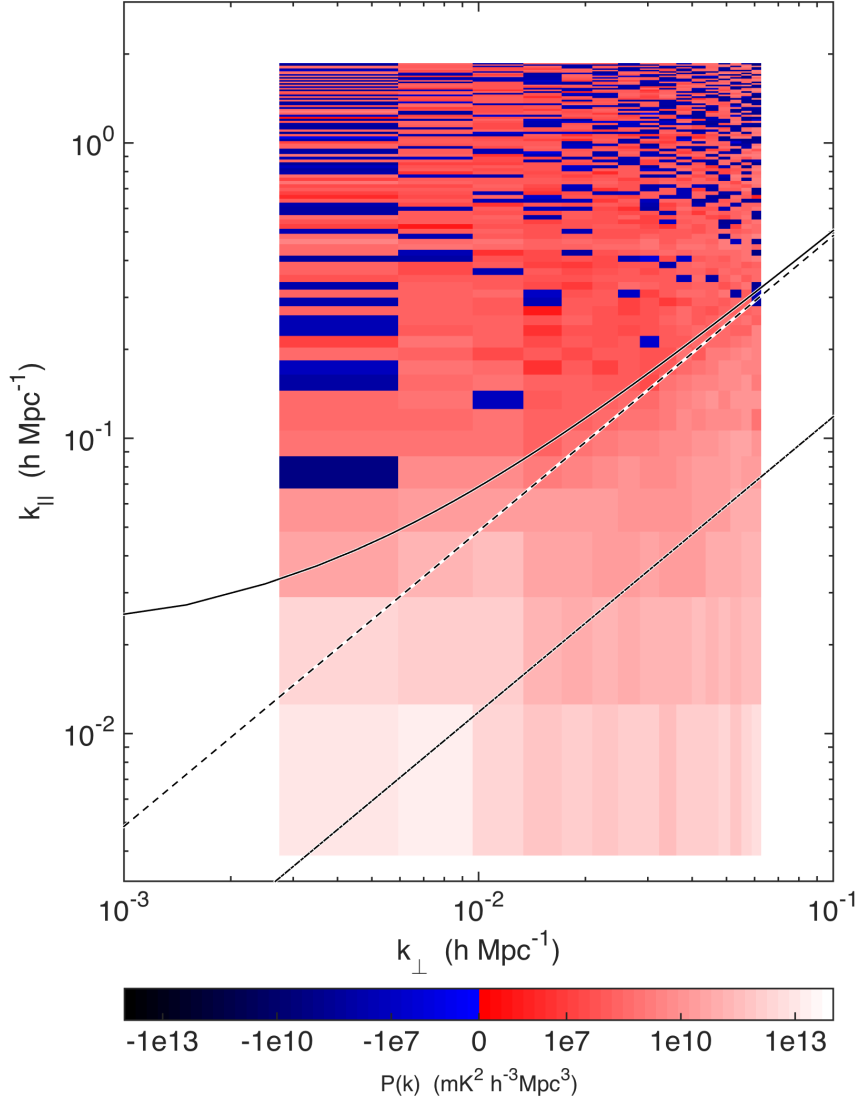
Figure 5-6: By using an estimator of the power spectrum with uncorrelated errors between bins, we can see that most of the EoR window is noise dominated in our power spectum measurement. Here we show the inverse hyperbolic sine of the power spectrum, which behaves linearly near zero and logarithmically at large magnitudes. Because we are taking a cross power spectrum between two data cubes with uncorrelated noise, noise dominated regions are equally likely to have positive power as negative power. Since we do not attempt to subtract a foreground bias, foreground contaminated regions show up as strongly positive. That includes the wedge, the bandpass line at $k_{\parallel} \approx 0.45\,h\,\mathrm{Mpc}^{-1}$ (see Figure 5-5), and some of the EoR window at low $k_{\perp}$ and relatively low $k_{\parallel}$, consistent with the suprahorizon emission seen in [182].

residual foregrounds and noise to the expected variance in each band power [120, 58, 59], which we plot in Figure 5-7. As we have chosen our power spectrum normalization $\mathbf{M}$ such that $\boldsymbol{\Sigma} \equiv \mathrm{Cov}(\widehat{\mathbf{p}})$ is diagonal, it is sufficient to plot the diagonal of $\boldsymbol{\Sigma}^{1/2}$, the standard deviation of each band power. The EoR window is seen clearly here as well. There is high variance at low and high $k_\perp$ where the uv coverage is poor, and also in the wedge due to foreground residuals. It is particularly pronounced in the bottom left corner, which is dominated by residual diffuse foregrounds.

As our error covariance represents the error due to both noise and foregrounds we expect to make in an estimate of the 21 cm signal, it is interesting to examine the "signal to error ratio" in Figure 5-8—the ratio of Figure 5-5 to Figure 5-7. The ratio is of order unity in noise dominated regions—though it is slightly lower than what we might naively expect due to our conservative estimate of $\boldsymbol{\Sigma}$ [59]. That explains the number of modes with very small values in Figure 5-7. In the wedge and just above it, however, the missubtracted foreground bias is clear, appearing as a high significance "detection" of the foreground wedge in the residual foregrounds. The bandpass miscalibration line at $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$ also appears clearly due to both foreground bias and possibly an underestimation of the errors. Hedging against this concern, we simply project out this line from our estimator that bins 2D power spectra into 1D power spectra by setting the variance of those bins to infinity.

Though useful for the careful evaluation of our techniques and of the instrument, the large bandwidth data cubes used to make Figures 5-5 and 5-6 encompass long periods of cosmic time over which the 21 cm power spectrum is expected to evolve. The cutoff is usually taken to be $\Delta z \lesssim 0.5$ [135]. These large data cubes also violate the assumption in [58] that channels of equal width in frequency correspond to equal comoving distances, justifying the use of the fast Fourier transform. Therefore, we break the full bandwidth into three 10.24 MHz segments before forming spherically averaged power spectra, and estimate the foreground residual covariance and power spectrum independently from each. We bin our 2D power spectra into 1D power spectra using the optimal estimator formalism of [59]. In our case, since we have chosen $\mathbf{M}$ such that $\boldsymbol{\Sigma}$ is diagonal, this reduces to simple inverse variance weighting
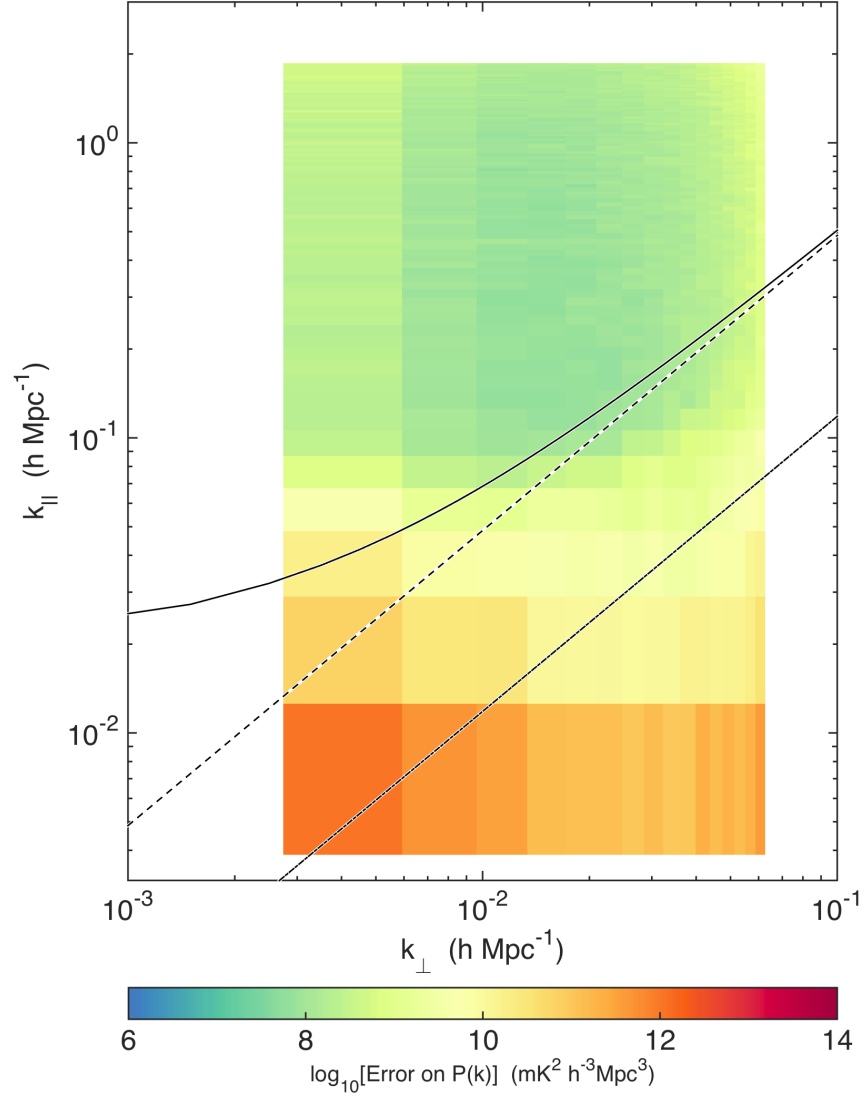
Figure 5-7: By including both residual foregrounds and noise in $\mathbf{C}$, our model for the covariance, we can calculate the expected variance on each band power in $\widehat{\mathbf{p}}$, which we show here. We see more variance at high (and also very low) $k_\perp$ where we have few baselines. We also see high variance at low $k_\parallel$ consistent with foregrounds. We see the strongest foregrounds at low $k_\perp$, which implies that the residual foregrounds have a very strong diffuse component that we have much to gain from better diffuse models to subtract. We also see that foreground-associated variance extends to higher $k_\parallel$ at high $k_\perp$, which is exactly the expected effect from the wedge. Both these observations are consistent with the structure of the eigenmodes we saw in Figure 5-1. Because we have chosen a normalization of $\widehat{\mathbf{p}}$ such that the $\mathrm{Cov}(\widehat{\mathbf{p}})$ is diagonal, this is a complete description of our errors. Furthermore, it means that the band powers form a mutually exclusive and collectively exhaustive set of measurements.
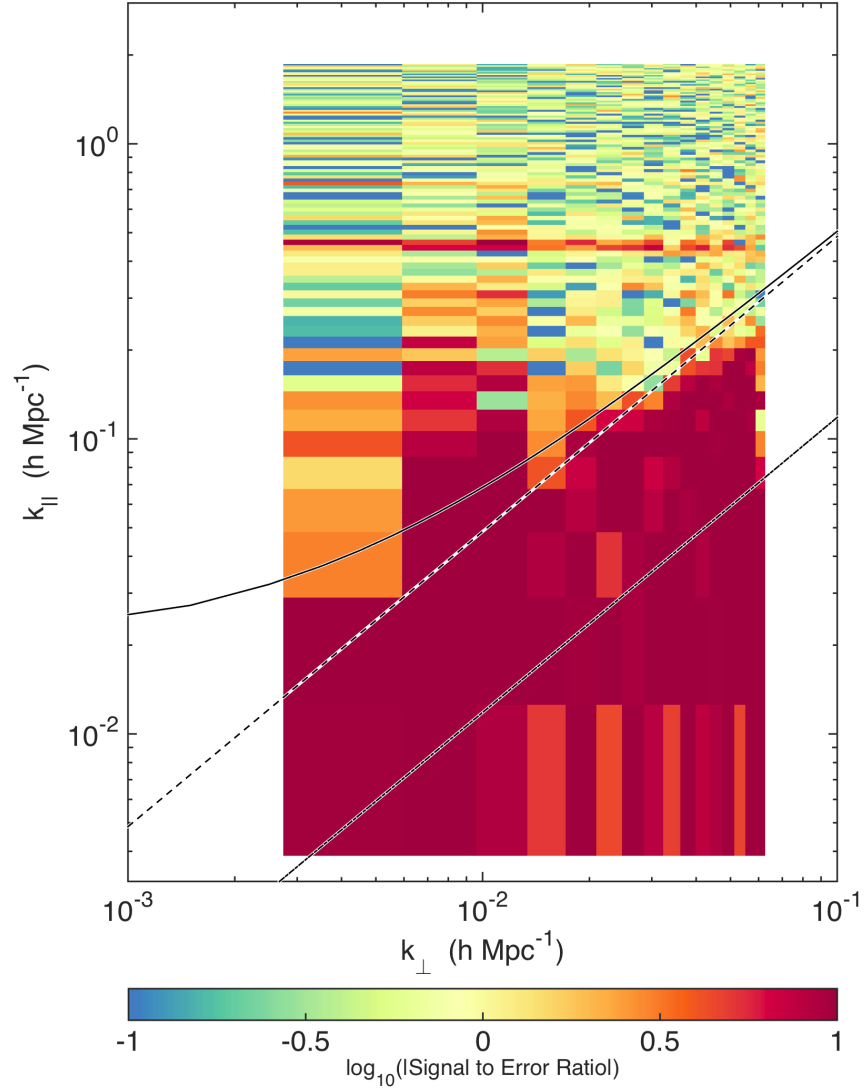
Figure 5-8: The foregrounds' wedge structure is particularly clear when looking at the ratio of our measured power spectrum to the modeled variance, shown here. Though the variance in foreground residual dominated parts of the $k_\perp$-$k_\parallel$ plane are elevated (see Figure 5-7), we still expect regions with signal to error ratios greater than one. This is largely due to the fact that we choose not to subtract a foreground bias for fear of signal loss. This figure shows us most clearly where the foregrounds are important and, as with Figure 5-6, it shows where we can hope to do better with more integration time and where we need better calibration and foreground modeling to further integrate down.

with the variance on modes outside the EoR window or in the $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$ line set to infinity.

In Figure 5-9 we show the result of that calculation as a "dimensionless" power spectra $\Delta^2(k) \equiv k^3 P(k)/2\pi^2$. We choose our binning such that the window functions (calculated as in [59] from our covariance model) were slightly overlapping.

Our results are largely consistent with noise. Since noise is independent of $k_\parallel$ and $k \approx k_\parallel$ for most modes we measure, the noise in $\Delta^2(k)$ scales as k^3. We see deviations from that trend at low k where modes are dominated by residual foreground emission beyond the horizon wedge and thus show elevated variance and bias in comparison to modes at higher k. Since we do not subtract a bias, even these "detections" are upper limits on the cosmological signal.

A number of barely significant "detections" are observed at higher k. Though we excise bins associated with the $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$ line, the slight detections may be due to leakage from that line. At higher z, the feature may due to reflections from cables of a different length, though some may be plausibly attributable to noise. Deeper integration is required to investigate further.

Our best upper limit at 95% confidence is $\Delta^2(k) < 3.7 \times 10^4\,\mathrm{mK}^2$ at $k = 0.18\,h\,\mathrm{Mpc}^{-1}$ around $z = 6.8$. Our absolute lowest limit is about 2 times lower than the best limit in [59], though the latter was obtained at substantially higher redshift and lower k, making the two somewhat incomparable. Our best limit is roughly 3 orders of magnitude better than the best limit of [59] over the same redshift range, and the overall noise level (as measured by the part of the power spectrum that scales as k^3) is more than 2 orders of magnitude smaller. This cannot be explained by more antenna tiles alone; it is likely that the noise level was overestimated in [59] due to insufficiently rapid time interleaving of the data cubes used to infer the overall noise level.

Although one cannot directly compare limits at different values of k and z, our limit is similar to the GMRT limit [167], $6.2 \times 10^4\,\mathrm{mK}^2$ at $k = 0.50\,h\,\mathrm{Mpc}^{-1}$ and $z = 8.6$ with $40\,\mathrm{h}$ of observation, and remains higher than the best PAPER limit [6] of $502\,\mathrm{mK}^2$ between $k = 0.15\,h\,\mathrm{Mpc}^{-1}$ and $k = 0.50\,h\,\mathrm{Mpc}^{-1}$ and $z = 8.4$ with 4.5 months of observation.
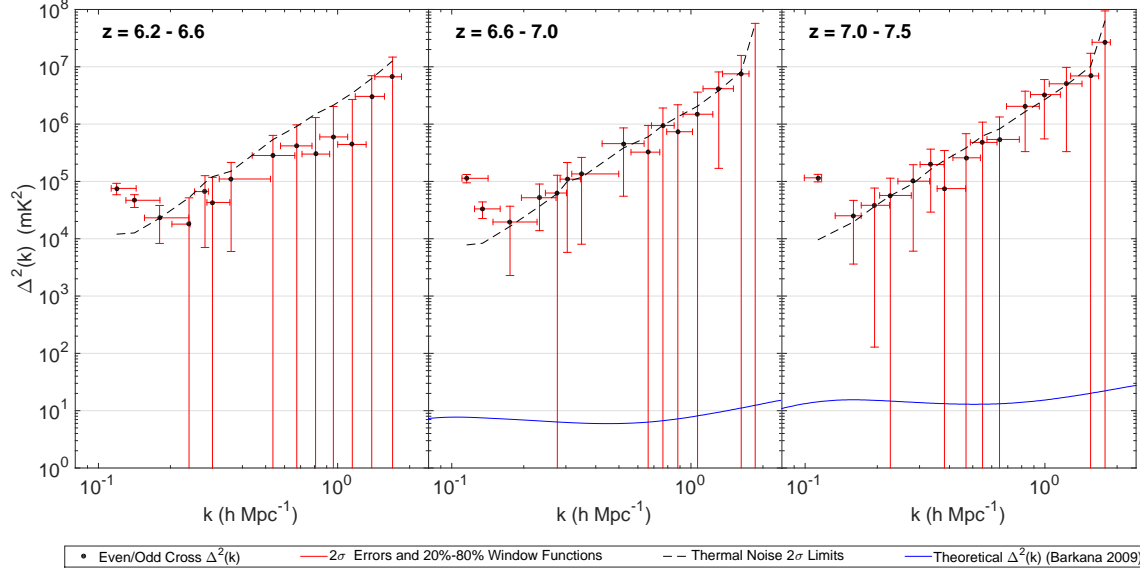
Figure 5-9: Finally, we can set confident limits on the 21 cm power spectrum at three redshifts by splitting our simultaneous bandwidth into three 10.24 MHz data cubes. The lowest k bins show the strongest "detections," though they are attributable to suprahorizon emission [182] that we expect to appear because we only cut out the wedge and a small buffer $(0.02\,h\,\mathrm{Mpc}^{-1})$ past it. We also see marginal "detections" at higher k which are likely due to subtle bandpass calibration effects like cable reflections. The largest such error, which occurs at bins around $k_\parallel \sim 0.45\,h\,\mathrm{Mpc}^{-1}$ and can be seen most clearly in Figure 5-8, has been flagged and removed from all three of these plots. Our absolute lowest limit requires $\Delta^2(k) < 3.7 \times 10^4\,\mathrm{mK}^2$ at 95% confidence at comoving scale $k = 0.18\,h\,\mathrm{Mpc}^{-1}$ and $z = 6.8$, which is consistent with published limits [167, 59, 173, 105, 6]. We also include a simplistic thermal noise calculation (dashed line), based on our observed system temperature. Though it is not directly comparable to our measurements, since it has different window functions, it does show that most of our measurements are consistent with thermal noise. For comparison, we also show the theoretical model of [10] (which predicts that reionization ends before $z = 6.4$) at the central redshift of each bin. While we are still orders of magnitude away from the fiducial model, recall that the noise in the power spectrum scales inversely with the integration time, not the square root.

In Figure 5-9 we also plot a theoretical model from [10] predicting that reionization has ended by the lowest redshift bin we measure. We remain more than 3 orders of magnitude (in mK^2) from being able to detect that particular reionization model, naively indicating that roughly 3000 h of data are required for its detection. This appears much larger than what previous sensitivity estimates have predicted for the MWA (e.g. [15]) in the case of idealized foreground subtraction.

However, much of this variance is due to the residual foregrounds and systematics in the EoR window identified by our empirical covariance modeling method, not thermal noise (see Figure 5-7). More integration will not improve those modes unless it allows for a better understanding of our instrument, better calibration, and better foreground models—especially of diffuse emission which might contaminate the highly sensitive bottom left corner of the EoR window. Eliminating this apparent "suprahorizon" emission, seen most clearly as detections in Figure 5-8 below $k \approx 0.2 \, h \, \mathrm{Mpc}^{-1}$, is essential to achieving the forecast sensitivity of the MWA [15]. If we can do so, we may still be able to detect the EoR with 1000 h or fewer. This is especially true if we can improve the subtraction of foregrounds to the point where we can work within the wedge, which can vastly increase the sensitivity of the instrument [15, 184]. On the other hand, more data may reveal more systematics lurking beneath the noise which could further diminish our sensitivity.

5.4 Summary and Future Directions

In this work, we developed and demonstrated a method for empirically deriving the covariance of residual foreground contamination, $\mathbf{C}^{FG}$, in observations designed to measure the 21 cm cosmological signal. Understanding the statistics of residual foregrounds allows us to use the quadratic estimator formalism to quantify the error associated with missubtracted foregrounds and their leakage into the rest of the EoR window. Because of the complicated interaction between the instrument and the foregrounds, we know that the residual foregrounds will have complicated spectral structure, especially if the instrument is not perfectly calibrated. By deriving our

model for $\mathbf{C}^{FG}$ empirically, we could capture those effects faithfully and thus mitigate the effects of foregrounds in our measurement (subject to certain caveats which we recounted in Section 5.2.4).

Our strategy originated from the assumption that the frequency-frequency covariance, modeled as a function of $|u|$, is the most important component of the foreground residual covariance. We therefore used sample covariances taken in annuli in Fourier space as the starting point of our covariance model. These models were adjusted to avoid double counting the noise variance and filtered in Fourier space to minimize the effect of noise in the empirically estimated covariances. Put another way, we combined our prior beliefs about the structure of the residual foregrounds with their observed statistics in order to build our models.

We demonstrated this strategy through the power spectrum analysis of a $3\,\mathrm{h}$ preliminary MWA data set. We saw the expected wedge structure in both our power spectra and our variances. We saw that most of the EoR window was consistent with noise, and we understand why residual foregrounds and systematics affect the regions that they do. We were also able to set new MWA limits on the 21 cm power spectrum from $z = 6.2$ to 7.5, with an absolute best 95% confidence limit of $\Delta^2(k) < 3.7 \times 10^4\,\mathrm{mK}^2$ at $k = 0.18\,h\,\mathrm{Mpc}^{-1}$ and $z = 6.8$, consistent with published limits [173, 105].

This work suggests a number of avenues for future research. Of course, improved calibration and mapmaking fidelity—especially better maps of diffuse Galactic structure—will improve power spectrum estimates and and allow deeper integrations without running up against foregrounds or systematics. Relaxing some of the mapmaking and power spectrum assumptions discussed in Section 5.2.4 may further mitigate these effects. A starting point is to integrate the mapmaking and statistical techniques of [61] with the fast algorithms of [58]. The present work is based on the idea that it is simpler to estimate $\mathbf{C}^{FG}$ from the data than from models of the instrument and the foregrounds. But if we can eliminate systematics to the point where we really understand $\mathbf{P}$, the relationship between the true sky and our dirty maps, then perhaps we can refocus our residual foreground covariance modeling effort on the statistics of the true sky residuals using the fact that $\mathbf{C}^{FG} = \mathbf{P}\mathbf{C}^{FG,\mathrm{true}}\mathbf{P}^{\mathsf{T}}$.

Obtaining such a complete understanding of the instrument will be challenging, but it may be the most rigorous way to quantify the errors introduced by missubtracted foregrounds and thus to confidently detect the 21 cm power spectrum from the epoch of reionization.

Chapter 6

MITEoR: A Scalable Interferometer for Precision 21 cm Cosmology

The content of this chapter was submitted the Monthly Notices of the Royal Astronomical Society *on June 12, 2014 and published [248] as* MITEoR: a scalable interferometer for precision 21 cm cosmology *on October 8, 2014.*

6.1   Introduction

Mapping neutral hydrogen throughout our universe via its redshifted 21 cm line offers a unique opportunity to probe the cosmic "dark ages," the formation of the first luminous objects, and the epoch of reionization (EoR). A suitably designed instrument with a tenth of a square kilometer of collecting area will allow tight constraints on the timing and duration of reionization and the astrophysical processes that drove it [184]. Moreover, because it can map a much larger comoving volume of our universe, it has the potential to overtake the Cosmic Microwave Background (CMB) as our most sensitive cosmological probe of inflation, dark matter, dark energy, and neutrino masses. For example [135], a radio array with a square kilometer of collecting area, maximal sky coverage, and good foreground maps could improve the sensitivity to spatial curvature and neutrino masses by up to two orders of magnitude, to $\Delta\Omega_k \approx 0.0002$ and $\Delta m_\nu \approx 0.007$ eV, and shed new light on the early universe by a 4σ

detection of the spectral index running predicted by the simplest inflation models favored by the BICEP2 experiment [3].

Unfortunately, the cosmological 21 cm signal is so faint that none of the current experiments around the world (LOFAR [107], MWA [220], PAPER [171], 21CMA [239], GMRT [166]) have detected it yet, although increasingly stringent upper limits have recently been placed [167, 59, 173]. A second challenge is that foreground contamination from our galaxy and extragalactic sources is perhaps four orders of magnitude larger than the cosmological hydrogen signal [54]. Any attempt to accurately clean it out from the data requires even greater sensitivity as well as more accurate calibration and beam modeling than the current state-of-the-art in radio astronomy (see Furlanetto et al. [74], Morales and Wyithe [154] for reviews).

Large sensitivity requires large collecting area. Since steerable single dish radio telescopes become prohibitively expensive beyond a certain size, the aforementioned experiments have all opted for interferometry, combining N (generally a large number) independent antenna elements which are (except for GMRT) individually more affordable. The LOFAR, MWA, PAPER, 21CMA and GMRT experiments currently have comparable N. The problem with scaling interferometers to high N is that all of these experiments use standard hardware cross-correlators whose cost grows quadratically with N, since they need to correlate all $N(N-1)/2 \sim N^2/2$ pairs of antenna elements. This cost is reasonable for the current scale $N \sim 10^2$, but will completely dominate the cost for $N \gtrsim 10^3$, making precision cosmology arrays with $N \sim 10^6$ as discussed in Mao et al. [135] infeasible in the near future, which has motivated novel correlator approaches such as Morales [152].

For the particular application of 21 cm cosmology, however, designs with better cost scaling are possible, as described in Tegmark and Zaldarriaga [210, 211]: by arranging the antennas in a hierarchical rectangular or hexagonal grid and performing the correlations using Fast Fourier Transforms (FFTs), thereby cutting the cost scaling to $N \log N$. This is particularly attractive for science applications requiring exquisite sensitivity at vastly different angular scales, such as 21 cm cosmology (where

short baselines are needed to probe the cosmological signal[1] and long baselines are needed for point source removal). Such hierarchical grids thus combine the angular resolution advantage of traditional array layouts with the cost advantage of a rectangular Fast Fourier Transform Telescope. If the antennas have a broad spectral response as well and their signals are digitized with high bandwidth, the cosmological neutral hydrogen gets simultaneously imaged in a vast 3D volume covering both much of the sky and also a vast range of distances (corresponding to different redshifts, *i.e.*, different observed frequencies.) Such low-cost arrays have been called *omniscopes* [210, 211] for their wide field of view and broad spectral range.

Of course, producing such scientifically rich maps with any interferometer depends crucially on our ability to precisely calibrate the instrument, so that we can truly understand how our measurements relate to the sky. Traditional radio telescopes rely on a well-sampled Fourier plane to perform self-calibration using the positions and fluxes of a number of bright point sources. At first blush, one might think that any highly-redundant array would be at a disadvantage in its attempt to calibrate the gains and phases of individual antennas. However, we can use the fact that redundant baselines should measure the same Fourier component of the sky to improve the calibration of the array dramatically and quantifiably. In fact, we find that the ease and precision of redundant baseline calibration is a strong rationale for building a highly-redundant array, in addition to the improvements in sensitivity and correlator speed.

Redundant calibration is useful both for current generation redundant arrays like MITEoR and PAPER and for future large arrays that will need redundancy to cut down correlator cost. Omniscopes must be calibrated in real time, because they

---

[1]It has been shown that the 21 cm signal-to-noise ratio (S/N) per resolution element in the *uv*-plane (Fourier plane) is $\ll 1$ for all current 21 cm cosmology experiments, and that their cosmological sensitivity therefore improves by moving their antennas closer together to focus on the center of the *uv*-plane and bringing its S/N closer to unity [151, 25, 139, 135, 118]. Error bars on the cosmological power spectrum have contributions from both noise and sample variance, and it is well-known that the total error bars on a given physical scale (for a fixed experimental cost) are minimized when both contributions are comparable, which happens when the $S/N \sim 1$ on that scale. This is why more compact 21 cm experiments have been advocated. This is also why early suborbital CMB experiments focused on small patches of sky to get $S/N \sim 1$ per pixel, and why galaxy redshift surveys target objects like luminous red galaxies that give $S/N \sim 1$ per 3D voxel.

do not compute and store the visibilities measured by each pair of antennas, but effectively gain their speed advantage by averaging redundant baselines in real time. Individual antennas therefore cannot be calibrated in post-processing. No calibration scheme used on existing low frequency radio interferometers has been demonstrated to meet the speed and precision requirements of omniscopes. Thus, the main goal of the MIT Epoch of Reionization experiment (MITEoR) and this paper is to demonstrate a successful redundant calibration pipeline that can overcome the calibration challenges faced by current and future generation instruments by performing automatic precision calibration in real time.

Building on past redundant baseline calibration methods by Wieringa [233] and others, some of us recently developed an algorithm which is both automatic and statistically unbiased, able to produce precision phase and gain calibration for all antennas in a hierarchical grid (up to a handful of degeneracies) without making any assumptions about the sky signal [124]. Once obtained, precision calibration solutions can in turn produce more accurate modeling of the synthesized and primary beams[2] [181], which has been shown to improve the quality of the foreground modeling and removal which is so crucial to 21 cm cosmology. It is therefore timely to develop a pathfinder instrument that tests how well the latest calibration ideas works in practice.

MITEoR is such a pathfinder instrument, designed to test redundant baseline calibration. We developed and successfully applied a real-time redundant calibration pipeline to data we took with our 64 dual-polarization antenna array during the summer of 2013 in The Forks, Maine. The goal of this paper is to describe the design of the MITEoR instrument, demonstrate the effectiveness of our redundant baseline calibration and absolute calibration pipelines, and use the calibration results to obtain an optimal scheme for estimating calibration parameters as a function of time and frequency.

This paper is organized as follows. We first describe in Section 6.2 the instru-

---

[2]For tile-based interferometers like the MWA and 21CMA, gain and phase errors in individual antennas (as opposed to tiles) do not typically get calibrated in the field, adding a fundamental uncertainty to the tile sky response.

ment, including the custom developed analog components, the 8 bit 128 antenna-polarization correlator, the deployment, and the observation history. In Section 6.3, we focus on precision calibration. We explain and quantitatively evaluate relative redundant calibration, and address the question of how often calibration coefficients should be updated. We also examine the absolute calibration, including breaking the degeneracies in relative calibration, mapping the primary beam, and measuring the array orientation. In Section 6.4, we summarize this work and discuss implications for future redundant arrays such as HERA [184].

6.2   The MITEoR Experiment

In theory, a very large omniscope can be built following the generalized architecture in Figure 6-1. On the other hand, it is crucial to demonstrate that automatic and precise calibration is possible in real-time using redundant baselines, since the calibration coefficients for each antenna must be updated frequently to allow the FFTs to combine the signals from the different antennas without introducing errors. In this section, we will present our partial implementation of this general design, including both the analog and the digital systems. Because the digital hardware is powerful enough to allow it, the MITEoR prototype correlates all 128 input channels with one another, rather than just a small sample as mentioned in the caption of Figure 6-1. This provides additional cross-checks that greatly aid technological development, where instrumentation may be particularly prone to systematics. This also allows us to explore the question of exactly how often and how finely in frequency we must measure visibilities to solve for calibration coefficients, a question we return to in Section 6.3. Since we chose to implement a full correlator, an additional FFT correlator would bring no extra information (simply computing the same redundant-baseline-averaged visibilities faster), so we leave the digital implementation of an FFT correlator to future work. In general, our mission is to empirically explore any challenges that are unique to a massively redundant interferometer array. Once these are known, one can reconfigure the cross-correlation hardware to perform spatial FFTs, thereby obtaining
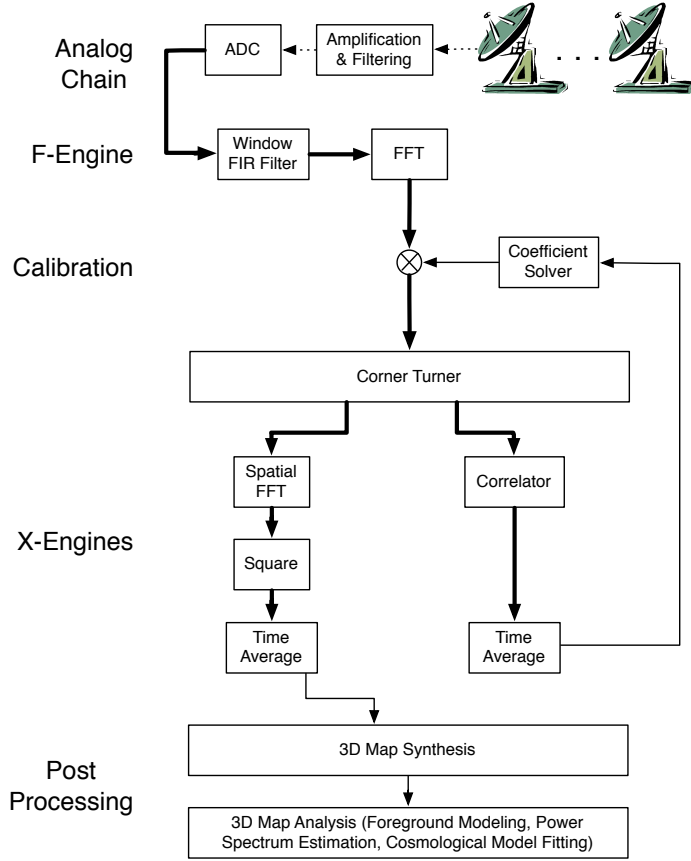
Figure 6-1: Data pipeline for a large omniscope that implements FFT correlator and redundant baseline calibration. First, a hierarchical grid of dual-polarization antennas converts the sky signal into volts, which get amplified and filtered by the analog chain, transported to a central location, and digitized every few nanoseconds. These high-volume digital signals (thick lines) get processed by field-programmable gate arrays (FPGAs) which perform a temporal Fourier transform. The FPGAs (or GPUs) then multiply by complex-valued calibration coefficients that depend on antenna, polarization and frequency, then spatially Fourier transform, square and accumulate the results, recording integrated sky snapshots every few seconds and thus reducing the data rate by a factor $\sim 10^9$. They also cross-correlate a small fraction of all antenna pairs, allowing the redundant baseline calibration software [124, 160] to update the calibration coefficients in real time and automatically monitor the quality of calibration solutions for instrumental malfunctions. Finally, software running on regular computers combine all snapshots of sufficient quality into a 3D sky ball or "data cube" representing the sky brightness as a function of angle and frequency in Stokes (I,Q,U,V) [211], and subsequent software accounts for foregrounds and measures power spectra and other cosmological observables.

an omniscope with $N \log N$ correlator scaling.

6.2.1 The Analog System

MITEoR contains 64 dual-polarization antennas, giving 128 signal channels in total. The signal picked up by the antennas is first amplified by two orders of magnitude in power by the low noise amplifiers (LNAs) built-in to the antennas. It is then phase switched in the swapper system, which greatly reduces cross-talk downstream. The signal is then amplified again by about five orders of magnitude in the line-drivers before being sent over 50 meter RG6 cables to the receivers. The receivers perform IQ demodulation on a desired 50 MHz band selected between 100 MHz and 200 MHz, producing two channels with adjacent 25 MHz bands, and sends the resulting signals into the digitization boards containing 256 analog-to-digital converters (ADCs) sampling at 50 MHz. The swappers, line-drivers and receivers we designed are shown in Figure 6-2.

When designing the components of this system, we chose to use commercially-available integrated circuits and filters whenever possible, to allow us to focus on system design and construction. In some cases (such as with the amplifiers) the cost of the IC is less than the cost of enough discrete transistors to implement even a rough approximation of the same functionality. Less expensive filters could be made from discrete components, but the characteristics of purchased modules tend to be better due to custom inductors and shielding. When we needed to produce our own boards as described below, our approach was to design, populate and test them in our laboratory, then have them affordably mass-produced for us by Burns Industries[3].

6.2.1.1 Antennas

The dual-polarization antennas used in MITEoR were originally developed for the Murchison Widefield Array [129, 220], and consist of two "bow-tie"-shaped arms as can be seen in Figure 6-8. They are inexpensive, easy to assemble, and sensitive to

---

[3] `http://www.burnsindustriesinc.com`
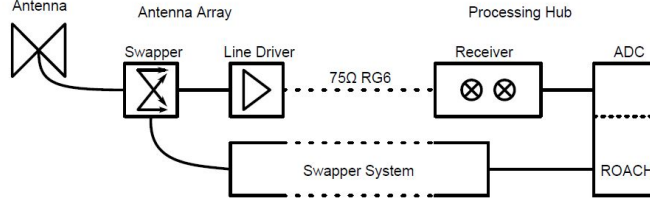
Figure 6-2: System diagram of the analog system. The signal received with an MWA "bow-tie" antenna is first amplified by the built-in low noise amplifier, then Walsh-modulated in the swapper module controlled by the swapper system. The signal is amplified again in the line driver and sent to the processing rack through 50 m long coaxial cables. In the processing rack, the signal first goes into the receiver, where it undergoes further amplification, frequency down-mixing and I/Q modulation from the 120-180 MHz range to the 0-25 MHz range. The analog chain ends with digitization on ADC connected to ROACH boards.

the entire band of our interest. The MWA antennas were designed for the frequency range 80-300 MHz, and have a built-in low noise amplifier with 20 dB of gain. The noise figure of the amplifier is 0.2 dB, and the 20 dB of gain means that subsequent gain stages do not contribute significantly to the noise figure[4].

6.2.1.2 Swappers (Phase Switches)

As with many other interferometers, crosstalk within the receivers, ADCs, and cabling significantly affects signal quality. We observe the cross-talk to depend strongly on the physical proximity of channel pairs, reaching as high as about −30 dB between nearest neighbor receiver channels. Our swapper system is designed to cancel out crosstalk during the correlator's time averaging by selectively inverting analog signals using Walsh modulation [193]. The signal from each antenna-polarization is inverted 50% of the time according to its own Walsh function, by an analog ZMAS-1 phase switch from Mini-Circuits located before the second amplification stage (line-driver), then appropriately re-inverted after digitization[5]. We perform the inversion once every millisecond, which is much longer than the ADC's 20 ns sample time, and much

---

[4]In a multi-stage amplifier, the contribution of each stage's noise figure is suppressed by a factor that is equal to the total gain of previous stages.

[5]Since the undesirable crosstalk signal is demodulated with a different Walsh function than it is modulated with, it will be averaged out due to orthogonality of Walsh functions.
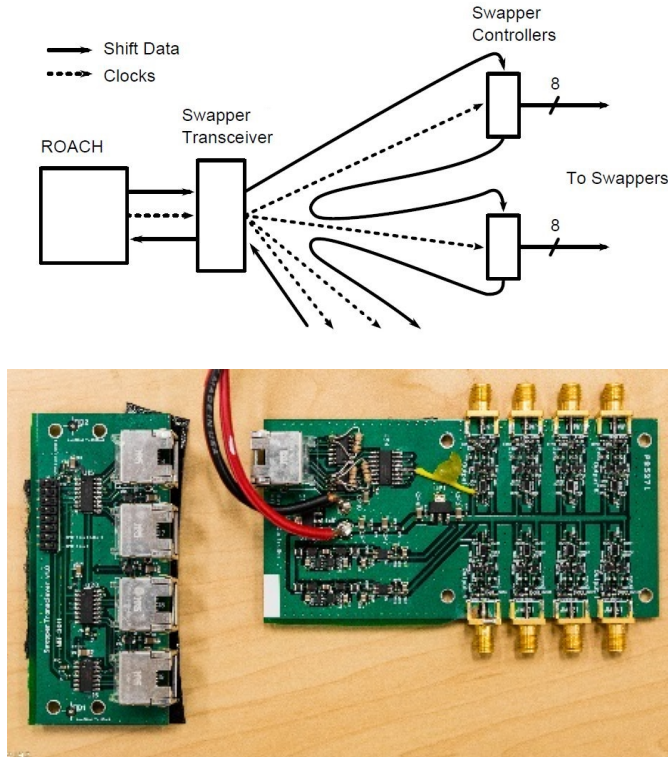
Figure 6-3: System diagram of our swapper signal system and physical components of the swapper transceiver (lower left) and swapper controller (lower right). The swapper is designed to reduce crosstalk between neighboring channels.

shorter than the averaging time of a few seconds[6]. This eliminates all crosstalk to first order [193]. If crosstalk reduction were the only concern, the ideal position for the swapper would be immediately after the antenna, in order to cancel as much crosstalk as possible. In practice, the swapper introduces a loss of about 3 dB, so we perform the modulation after the LNA to avoid adding noise (raising the system temperature). To evaluate the effectiveness of the swapper modules, we sent a monotone signal into one single channel of the receivers while leaving other channels open, and measured the correlation between the signal channel and each empty channels with the swapper turned on and off. We then repeated this while varying the signal frequency over the full range of interest. As seen in Figure 6-4, the swapper system attenuates crosstalk in the receiver and ADC by as much as 50 dB over the frequency band of interest,

---

[6]The inversion cannot be too frequent, because we need to discard data during the analog inversion process which takes a few microseconds. At the same time, the inversion needs to be frequent enough to average out the cross-talk.
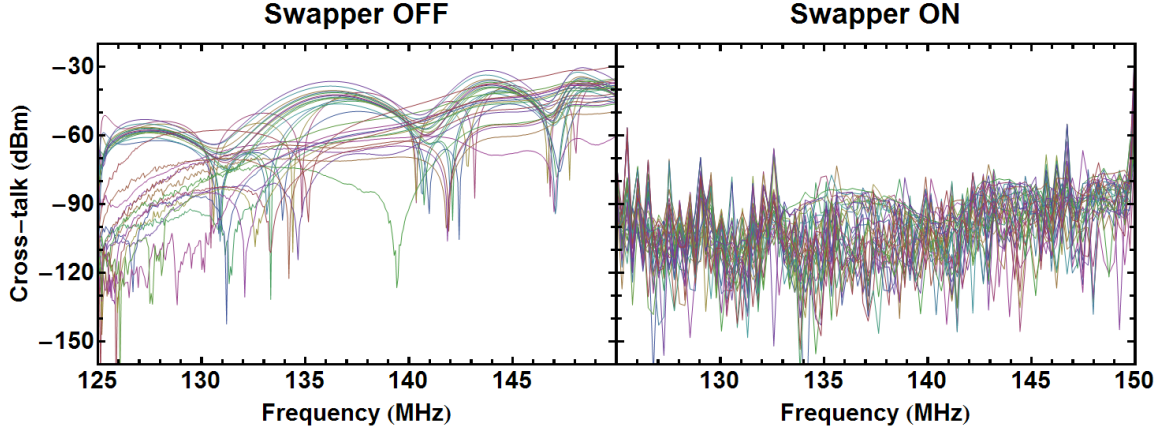
Figure 6-4: Plots of cross-talk power measured in the laboratory. The swapper suppresses crosstalk between channels by as much as 50 dB. To measure these curves we fed a 0 dBm sinusoidal signal into input channel 0 of the receivers and left the other 31 input channels open. We then measured the correlations between channel 0 and all 31 empty channels, due to crosstalk from channel 0. We repeated the procedure with input frequencies from 125-150 MHz and obtained the results shown above.

typically reducing it to being of order -80 dB for strongly afflicted signal pairs.

6.2.1.3 Line-Driver

A line-driver (Figure 6-5) amplifies a single antenna's signal from one of its two polarization channels while also powering its LNA. Line-drivers only handle a single channel to reduce potential crosstalk from sharing a printed circuit board. They are placed within a few meters of the antennas in order to reduce resistive losses from powering the antenna at low voltage. Additional gain that they provide early in the analog chain helps the signal overpower any noise picked up along the way to the processing hub, and maintains the low noise figure set up by the LNA. To further reduce potential radio-frequency interference (RFI), we chose to power the line-drivers with 58Ah 6V sealed lead acid rechargeable batteries during the final 64-antenna deployment, rather than 120 VAC to 6 VDC adapters (whose unwanted RF-emission may have caused occasional saturation problems during our earlier expeditions).

302
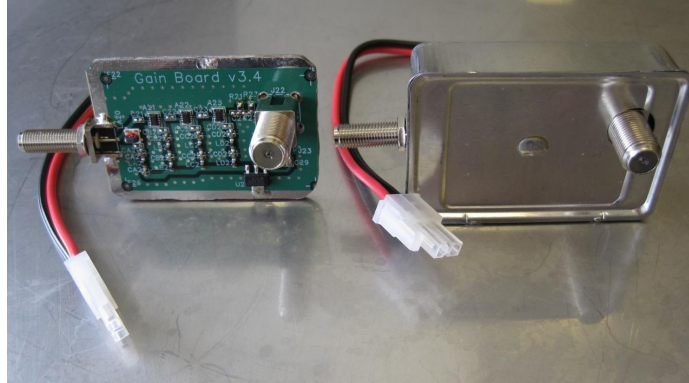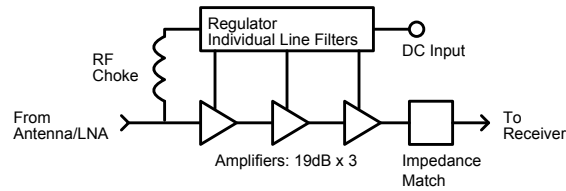
Figure 6-5: System diagram and physical components of the line drivers. The line driver we designed takes the signal in the 50Ω coaxial cable from the antenna LNA and amplifies it by 51 dB, in order to overpower noise picked up in the subsequent 75Ω coaxial cable and further processing steps up to 50 meters away. It operates on 5V DC and also provides DC bias power to the antenna's LNA through the 50 Ohm cable.

6.2.1.4 Receiver

Our receivers (Figure 6-6) take input from the line-drivers, bandpass filter the incoming signals, amplify their power level by 23 dB, and IQ-demodulate them. The resulting signals go directly to an ADC for digitization. Receivers are placed near the ADCs to which they are connected to reduce cabling for local oscillator (LO) distribution and ADC connections. IQ demodulation is used, which doubles received bandwidth for a given ADC frequency at the cost of using two ADC channels, and has the advantage of requiring only a single LO and low speed ADCs. The result is 40 MHz of usable bandwidth[7] anywhere in the range 110-190 MHz, with a 2-3 MHz gap centered around the LO frequency due to bandpass filters. The receiver boards have five pins allowing their signals to be attenuated by any amount between 0 dB and 31 dB (in steps of 1 dB) before the second amplification stage, to avoid saturation

---

[7]Due to limitations in our FPGAs' computing power, only 12.5 MHz of digitized data are correlated and stored at any instant.
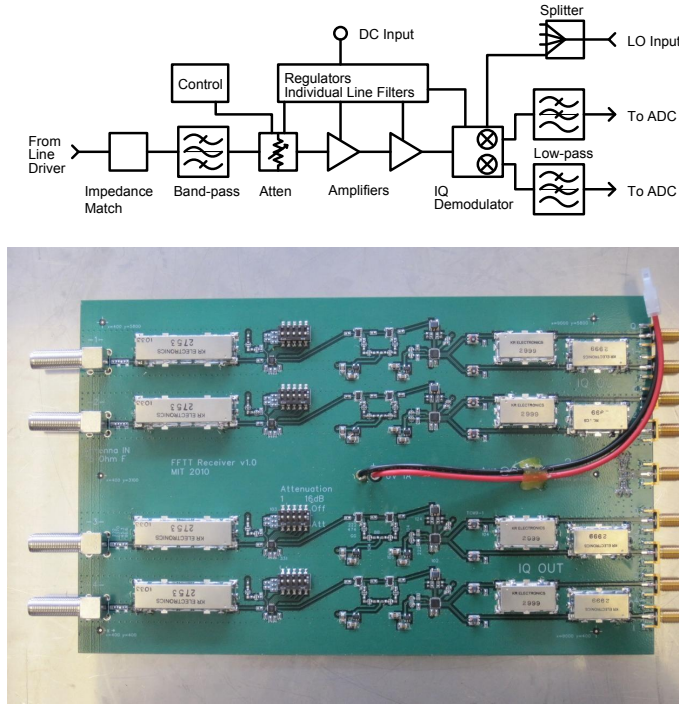
Figure 6-6: System diagram and physical component of the receiver boards. The boards take the signals arriving from four line drivers, band-pass filter and amplify them, then use a local oscillator to frequency shift them from the band of interest to a DC-centered signal suitable for input to the ADC.

and non-linearities from RFI and to attain signal levels optimal for digitization.

6.2.2 The Digital System

We designed MITEoR's digital system (Figure 6-7) to be highly compact and portable. The entire system occupies 2 shock-mounted equipment racks on wheels, each measuring about 1 m on all sides. It takes in data from 256 ADC channels (64 antennas with I and Q signals for polarizations), Fourier transforms each channel, reconstructs IQ demodulated channels back to 128 corresponding antenna channels, computes the cross-correlations of all pairs of the 128 antenna channels with 8 bit precision, and then time-averages these cross-correlations. Although standard 4 bit correlators suffice for most astronomical observation tasks, the better dynamic range of our 8 bit correlator allows us to observe faint astronomical signals at the same time as 10^3 times brighter ORBCOMM satellites, whose enormous signal-to-noise has proved in-

304

Figure 6-7: The entirety of our 128 antenna-polarization digital correlator system, packaged in two portable shock mounted racks. The two black chassis and two silver chassis in the middle of each rack are F-engines (ROACH) and X-engines (ROACH2), respectively. Above the ROACHes are 32 receiver boards that input the signals from 128 line drivers via F-cables. The blue lit area below the ROACHes contains various clocking devices responsible for synchronization, whereas the chassis below the ROACHes on the right hand side is the 8TB data acquisition server.

valuable in characterizing various aspects of the system (see Sections 6.3.2.2, 6.3.2.3, and 6.3.3). The digital hardware is capable of processing an instantaneous bandwidth of $12.5\,\mathrm{MHz}$ with $49\,\mathrm{kHz}$ frequency bins. It averages those correlations and then writes them to disk every few seconds (usually either 2.6 or 5.3 seconds).

While one of the advantages of a massively redundant interferometer array is the ability to reduce costs by performing a spatial FFT rather than a full cross-correlation, we have not implemented FFT correlation in the current MITEoR prototype as the hardware is powerful enough to correlate all antenna pairs in real time (the feasibility of implementing FFT correlation on the ROACH platform has been demonstrated by Foster et al. [68]). Rather, the goal of MITEoR is to quantify the accuracy that automatic redundant baseline calibration can attain, thereby experimentally characterizing all of the unknowns in the system, such as unexpected analog chain systematics and other barriers to finding good calibration solutions.

We adopted the widely-used F-X scheme in MITEoR's digital system. We have 4 synchronized F-engines that take in data from 4 synchronized 64-channel ADC

305

boards, which run at 12 bits and 50Ms/s. The F-engines perform the FFT and IQ reconstruction, and distribute the data onto 4 X-engines through 16 10GbE links. The 4 asynchronous X-engines each perform full correlation on 4 different frequency bands on all 128 antenna polarizations, and send the time averaged results to a computer for data storage.

To implement the computational steps of the MITEoR design, we used Field Programmable Gate Arrays (FPGAs). These devices can be programmed to function as dedicated pieces of computational hardware. Each F-engine and X-engine is implemented by one Xilinx FPGA (Virtex-5 for F-engines and Virtex-6 for X-engines). These FPGAs are seated on custom hardware boards developed by the CASPER collaboration[8] [169]. We also use the software tool flow developed by CASPER to design the digital system. The CASPER collaboration is dedicated to building open-source programmable hardware specifically for applications in astronomy. We currently use two of their newer devices, the ROACH[9] (Reconfigurable Open Architecture Computing Hardware) for the F-engines, and the ROACH 2[10] for the X-engines. The main benefit of using CASPER hardware is that it facilitates the time-consuming process of designing and building custom radio interferometry hardware. The CASPER collaboration also offers a large open-source library of FPGA blocks for commonly used signal processing structures such as polyphase filter banks, FIR filters and fast Fourier transform blocks [168]. However, due to MITEoR's ambitious architecture, involving both extreme compactness, an 8-bit correlator, and tight inter-ROACH synchronization constraints, we custom-designed most of the digital FPGA blocks. The specifications of our latest correlator are listed in Table 6.1.

6.2.3 MITEoR Deployment and Data Collection

We deployed MITEoR in The Forks, Maine, which our online research suggested might be one of the most radio quiet region in the United States at the frequencies

---

[8]https://casper.berkeley.edu/
[9]https://casper.berkeley.edu/wiki/ROACH
[10]https://casper.berkeley.edu/wiki/ROACH-2_Revision_2

Antenna	MWA dual-pol bow-tie
Antenna count	64×2 polarizations
Array configuration	8×8 grid
ADC	4×64-channel 50 Msps
F-engine	4 ROACHes with Virtex-5
X-engine	4 ROACH2s with Virtex-6
Correlator precision	8 bits
Frequency range	110-190 MHz
Instantaneous bandwidth	12.5 MHz (50 MHz digitized)
Frequency resolution	49 kHz
Time resolution	≥ 2.68 s

Table 6.1: List of MITEoR specifications. We observed with two different 8 by 8 array configurations, one with 3 m separation and one with 1.5 m separation. We observed ORBCOMM band with 2.68 s resolution, and we chose a resolution of 5.37 s for other bands.

of interest[11][24]. We deployed the first prototype in September 2010, and performed a successful suite of test observations with an 8-antenna interferometer. In May 2012, we completed and deployed a major upgrade of the digital system to fully correlate $N = 16$ dual-polarization antennas. With the experience of this successful deployment, we further upgraded the digital system to accommodate $N = 64$ dual-polarization antennas, which led to our latest deployment in July 2013 and the results we describe in this paper.

The MITEoR experiment was designed to be portable and easy to assemble. The entire experiment was loaded into a 17 foot U-Haul truck and driven to The Forks. It took a crew of 15 people less than 2 days to assemble the instrument and bring it to full capacity. A skeleton crew of 3 members stayed on site for monitoring and maintenance for the following two weeks, during which we collected more than 300 hours of data. Subsequently, a demolition crew of 5 members disassembled and packed up MITEoR in 6 hours and concluded the successful deployment.

During the deployment, we scanned through the frequency range 123.5-179.5 MHz,

---

[11]The Forks has also been successfully used to test the EDGES experiment [27], and we found the RFI spectrum to be significantly cleaner than at the National Radio Astronomy Observatory in Green Bank, West Virginia at the very low (100-200 MHz) frequency range that is our focus: the entire spectrum at The Forks is below -100 dBm except for one -89.5 dBm spike at 150MHz.

Figure 6-8: Part of the MITEoR array during the most recent deployment in the summer of 2013. 64 dual-polarization antennas were laid on a 21 m by 21 m regular grid with 3 m separation. The digital system was housed in the back of a shielded U-Haul truck (not shown).

with at least 24 consecutive hours at each frequency. We used two different array layouts for most of the frequencies we covered. The observation began with the antennas arranged in a regular 8 by 8 grid, with 3 meter spacing[12] between neighboring antennas, which we later reconfigured to an 8 by 8 regular grid with 1.5 meter spacing for a more compact layout (which provides better signal-to-noise ratio on the 21 cm signal). The total volume of binary data collected was 3.9TB, and in the rest of this paper, we demonstrate the results of our various calibration techniques using this data set.

6.3 Calibration Results

As we have emphasized above, the precision calibration of an interferometer is essential to its ability to detect the faint cosmological imprint upon the 21 cm signal, and the key focus of MITEoR is to determine how well real-time redundant calibration can be made to work in practice. In this section we describe the calibration scheme that we have designed and implemented and quantify its performance. We first constrain the *relative* calibration between antennas, utilizing both per-baseline algorithms and redundant-baseline calibration algorithms [124]. We then build on these relative

---

[12]We aligned the antenna positions using a laser-ranging total station, and measured their positions with millimeter level precision. The median deviation from a perfect grid is 2 mm in the N-S direction, 3 mm E-W, and 28 mm vertical, primarily caused by the fact that the deployment site had not been leveled.
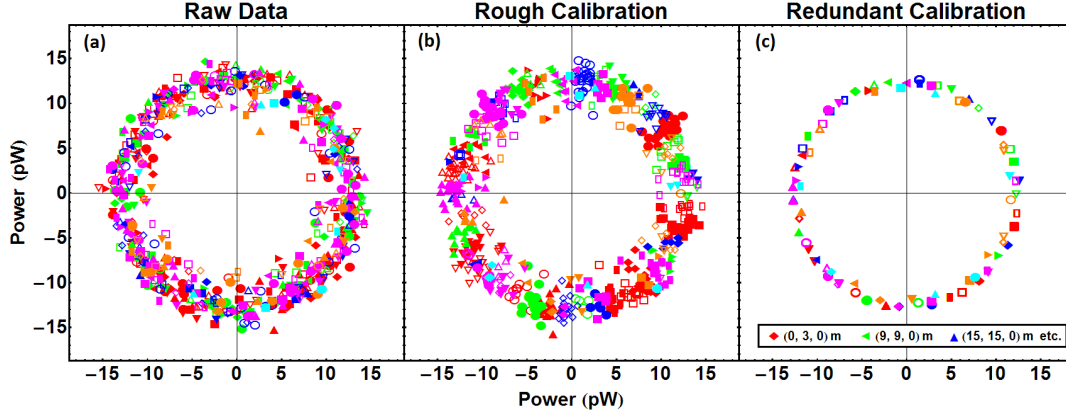
Figure 6-9: Illustration of three stages in the redundant baseline calibration pipeline. Each panel is a complex plane, and each point is a complex visibility for a specific pair of antennas at 137.1 MHz during the passage of an ORBCOMM satellite. Each unique combination of color and shape stands for one set of redundant baselines. In an ideal world, all identical symbols, such as all upright red triangles, should have the same value thus overlap exactly. Due to noise, they should cluster together around the same complex value. In panel (a) showing raw data, the redundant baselines have almost no clustering visible—for example, red filled circles can be found throughout the plot. After crude calibration in panel (b), we see most points falling into clustered segments—though the clustering is still far from exact. Finally in panel (c), after performing log calibration, we see that all points corresponding to each redundant baseline are almost exactly overlapping, with no visible deviation due to the high signal-to-noise. While the difference is not visible here, linear calibration can further improve log calibration results, as shown in Figure 6-11.

calibration results to constrain the *absolute* calibration of the instrument, including breaking the few degeneracies inherent to redundant calibration.

6.3.1 Relative Calibration

6.3.1.1 Overview

The goal of relative calibration is to calibrate out differences among antenna elements caused by non-identical analog components, such as variations in amplifier gains and cable lengths, which may be functions of time and frequency. We parametrize the calibration solution as a time- and frequency-dependent multiplicative complex gain g_i for each of the 128 antenna-polarizations. Calibrating the interferometer amounts to solving for the coefficients g_i and undoing their effects on the data. Our calibration

309

scheme revolves around calibration methods that heavily utilize the redundancy of our array, whose efficacy we aim to demonstrate with MITEoR. The current redundant calibration pipeline involves three steps, as illustrated in Figure 6-9:

1. Rough calibration computes approximate calibration phases using knowledge of the sky.

2. Logarithmic calibration ("logcal") decomposes roughly calibrated data into amplitudes and phases and computes least square fits for amplitude and phase separately.

3. Linear calibration ("lincal") takes the relatively precise but biased results from logcal and computes unbiased calibration parameters with even higher precision.

Although logcal and lincal have been previously proposed [233, 124, 136], they both fail in their original form if the phases of g_i are not close to 0.[13] In practice, the phases of g_i can be anywhere in the interval $[0, 2\pi)$. To overcome these practical challenges, we introduced various improvements to these algorithms. In the following sections, we describe our improvements to calibration algorithms in detail, and demonstrate the effectiveness of our calibration by obtaining $\chi^2/\text{DoF} \approx 1$ for the majority of our data. We then analyze these calibration parameters to construct a Wiener filter to optimally average them over time and frequency, which also tells us how frequently we need to calibrate in time and frequency.

6.3.1.2 Rough Calibration

The goal of rough calibration is to obtain reliable initial phase estimates for the calibration parameters to enable the subsequent more sophisticated algorithms. This step does not have to involve redundancy, thus it can be done with any standard calibration techniques, for example self-calibration [189, 30]. The rough calibration algorithm that we describe below is computationally cheap and can robustly improve upon raw data even when a few antennas have failed.

---

[13]Logcal requires phase calibrations close to 0 to avoid phase wrapping issues, whereas lincal requires phase calibrations close to 0 to converge.

At a given time and frequency, we have both the measured visibilities, v_{ij}, and v_{ij}^{model}, a rough model of the true sky signal[14], where indices i and j represent antenna number. We first compute the phase difference between each measured visibility and its prediction. We then pick one reference antenna and subtract the phases of its visibilities from the phases of other visibilities to obtain a list of estimated phase calibration for each antenna. Finally, we take the median of these calibration phases to obtain a robust phase calibration estimator for each antenna. More concretely, we use the following procedure:

1. Construct a matrix $\mathbf{M}$ of phase differences where $M_{ij} = -M_{ji} = \arg(v_{ij}/v_{ij}^{\text{model}})$.

2. Define the first antenna as the reference by subtracting the first column of $\mathbf{M}$ from all columns to obtain $M'_{jk} = M_{jk} - M_{j0}$.

3. Obtain rough phase calibration parameters $\phi_k \equiv \arg(g_k)$ by computing the median angle of column k in M', defined as

$$\phi_k \equiv \arg\left[\text{median}_j\{\exp(iM'_{jk})\}\right]$$
$$= \arg\left[\text{median}_j\{\cos(M'_{jk})\}\right.$$
$$\left. + i\,\text{median}_j\{\sin(M'_{jk})\}\right]. \tag{6.1}$$

For stable instruments, the true calibration parameters have very small variation over days, so we can use one set of rough calibration parameters from a single snapshot in time for data from all other times. Thus we pick a snapshot at noon when each v_{ij}^{model} can be easily computed from position of the Sun, and use the resulting raw calibration parameters as the starting point for logcal at all other times.

6.3.1.3   Log Calibration and Linear Calibration

To explain our redundant calibration procedure, we first need to briefly reintroduce the formalism developed in Liu et al. [124]. Suppose the ith antenna measures a signal

---

[14]Since we are trying to obtain an initial estimate, the model does not have to be very accurate.

s_i at a given instant. This signal can be expressed in terms of a complex gain factor g_i, the antenna's instrumental noise contribution n_i, and the true sky signal x_i that would be measured in the limit of perfect gain and no noise:

$$s_i = g_i x_i + n_i. \tag{6.2}$$

Under the standard assumption that the noise is uncorrelated with the signal, each baseline's measured visibility is the correlation between the two signals from the two antennas:

$$\begin{aligned} v_{ij} &\equiv \langle s_i^* s_j \rangle \\ &= g_i^* g_j \langle x_i^* x_j \rangle + g_i^* \langle x_i^* n_j \rangle + g_j \langle n_i^* x_j \rangle + \langle n_i^* n_j \rangle \\ &= g_i^* g_j y_{i-j} + n_{ij}^{\text{res}}, \end{aligned} \tag{6.3}$$

where we have denoted the true correlation $\langle x_i^* x_j \rangle$ by y_{i-j},[15] the noise from each antenna by n_i, the noise for each baseline by n_{ij}^{res}, and expectation values (effectively time averages) by angled brackets $\langle \dots \rangle$. In a maximally redundant array such as MITEoR, the number of unique baselines is much smaller than the total number of baselines. Therefore, we can treat all the g_is and the y_{i-j}s as unknowns while keeping the system of equations (6.3) overdetermined, enabling fits for both despite the presence of instrumental noise.

In Liu et al. [124], some of us proposed logcal and lincal, and we have implemented both for calibrating MITEoR data. In log calibration, we take the logarithm of both sides of Equation 6.3 and obtain a linearized equation in logarithmic space. We then perform a least squares fit for the system of equations

$$\log v_{ij} = \log g_i^* + \log g_j + \log y_{i-j}, \tag{6.4}$$

---

[15]Following Liu et al. [124], we use y_{i-j} instead of y_{ij} to emphasize that in a redundant array, the number of unique baseline visibilities can be much smaller than number of measured visibilities. The complete expression should be $y_{u(i,j)}$, where $u(i,j)$ means that baseline ij corresponds to the uth unique baseline.

where we solve for $\log g_i$ and $\log y_{i-j}$. Because the least squares fit takes place in log space whereas the noise is additive in linear space, the best fit results are biased. Linear calibration, on the other hand, is unbiased [124]. The lincal method performs a Taylor expansion of Equation 6.3 around initial estimates g_i^0 and y_{i-j}^0 and obtains a system of linearized equations

$$v_{ij} = g_i^{0*} g_j^0 y_{i-j}^0 + g_i^{1*} g_j^0 y_{i-j}^0 + g_i^{0*} g_j^1 y_{i-j}^0 + g_i^{0*} g_j^0 y_{i-j}^1, \qquad (6.5)$$

where we solve for g_i^1 and y_{i-j}^1. For a detailed description of the logcal and lincal algorithms and their noise properties, we direct the reader to Liu et al. [124], Marthi and Chengalur [136]. We now describe some essential improvements to these algorithms.

Logcal was first thought to be unable to calibrate the phase component due to phase wrapping, since logcal has no way to recognize that $0°$ and $360°$ are the same quantity. Consider, for example a pair of redundant baselines that measure phases of $0.1°$ and $359.9°$ respectively. We can infer that they each only need a very small phase correction ($\pm 0.1°$) to agree perfectly. However, since logcal treats the difference between them as $359.8°$ rather than $0.2°$, it will calibrate by averaging $0.1°$ and $359.9°$ to $180°$, which is completely wrong.

We made two improvements to the logcal method to guard against this. The first is to perform rough calibration beforehand, as described in Section 6.3.1.2. The second is to re-wrap the phases of v_{ij}. While rough calibration can make the phase errors relatively small[16], that improvement alone is not sufficient, since $0°$ and $360°$ are still treated as different quantities. Thus we need to intelligently wrap the phases of the input vector before feeding it into logcal. This is done in two simple steps. For a snapshot of rough calibrated visibilities at given time and frequency, v_{ij}, we first estimate the true phases of each group of redundant baselines, $\arg(y_{i-j})$, by computing median angles of measured phases using Eq. 6.1. Then for each measured phase, we add or subtract 2π until it is within $\pm\pi$ of $\arg(y_{i-j})$. This eliminates the phase wrapping problem.

---

[16]In our experience, they need to be less than about 20 degrees to ensure that the subsequent calibration steps converge reliably.

Unlike logcal, lincal is an unbiased algorithm, but it relies on a set of initial estimates for the correct calibration solutions to start with. The output of lincal can be fed back into the algorithm and it can iteratively improve upon its own solution. However, the algorithm converges to the right answer only if the initial estimates are good. In practice, we find that three iterations of lincal typically produces excellent convergence, because the outputs of logcal are already decent estimates of the calibration solutions. Thus, by improving logcal, we also greatly improve lincal's effectiveness.

Our current calibration pipeline performs all steps of redundant calibration in less than 1 millisecond on a single processor core for a data slice at one time and one frequency channel, which is an order of magnitude faster than the rate data is saved onto disk. It is carried out by our open source `Omnical` package, coded in C++/Python.[17] Thus there should be no computational challenge in performing the above described calibration procedure in real-time for any array with less than 10^3 elements. For a future omniscope that has as many as 10^6 elements, there are two ways to reduce the computational cost. The first is to calibrate less frequently in time and frequency, and we will discuss in detail the minimal sampling frequency in Section 6.3.1.5. The other is to adapt a hierarchical redundant calibration scheme, where instead of calibrating all visibilities at the same time, one can calibrate the array in a hierarchical fashion whose computational cost scales only linearly with the number of elements. We discuss more details regarding hierarchical redundant calibration in Appendix 6.B.

6.3.1.4 χ^2 and Quality of Calibration

One of the many advantages of redundant calibration is it allows for the calculation of a χ^2 for every snapshot to quantify how accurate the estimated visibilities are for each unique baseline, even without any knowledge of the sky. For a set of visibilities at a given time and frequency, v_{ij}, with calibration results g_i and y_{i-j}, we define χ^2

---

[17]The package supports the miriad file format and is easily adapted to work with other file formats. To obtain a copy, please contact jeff_z@mit.edu.
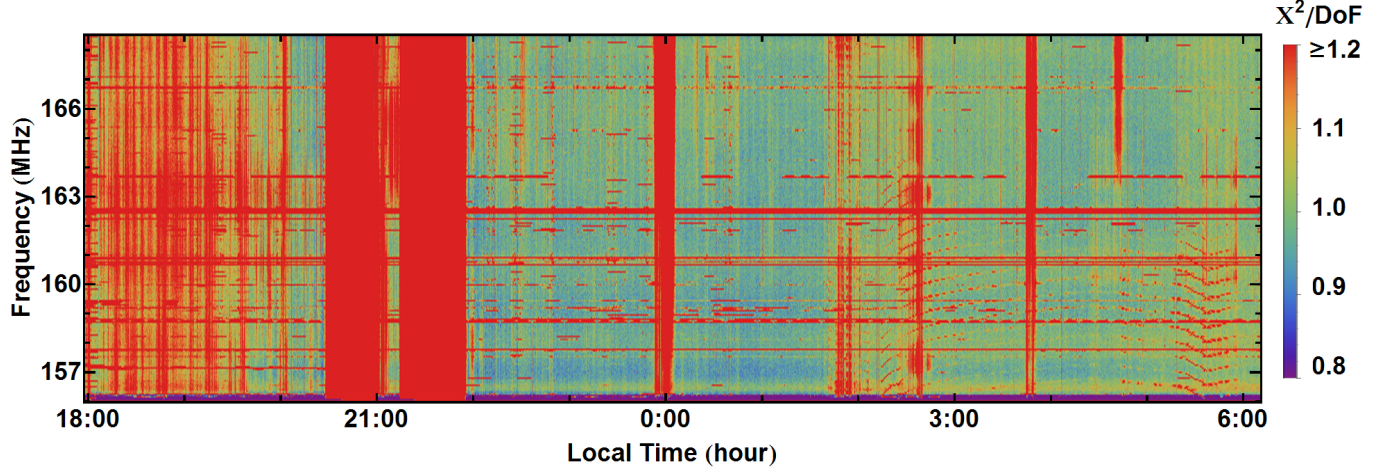
Figure 6-10: Waterfall plot of χ^2/DoF for a day's data. This demonstrates the stability of our instrument as well as the effectiveness of using χ^2/DoF as a indicator of data quality. We evaluate χ^2/DoF every 5.3 seconds and every 49 kHz. For the majority of the night time data, χ^2/DoF is close to 1. We flag all data with χ^2 larger than 1.2, which are marked red in this plot and account for 20% of this data set. The amount of detailed structure in the flagged area (around 18:00 for example) shows the χ^2 flagging technique's sensitivity to rapidly changing data quality.

as

$$\chi^2 = \sum_{ij} \frac{|v_{ij} - y_{i-j}g_i^*g_j|^2}{\sigma_{ij}^2},$$ (6.6)

where σ_{ij}^2 is the noise contribution to the variance of the visibility v_{ij}. The effective number of degrees of freedom (DoF) is

$$\text{DoF} = N_{\text{measurements}} - N_{\text{parameters}}$$

$$= N_{\text{baselines}} - (N_{\text{antennas}} + N_{\text{unique baselines}}).$$ (6.7)

The numerator in Equation 6.6 represents the deviation of measured data, v_{ij}, from the best fit redundant model, $y_{i-j}g_i^*g_j$. Thus χ^2/DoF can be interpreted as the non-redundancy in measured data divided by the expected non-redundancy from pure noise. If the data agrees perfectly with the redundant model (with noise) and is free from systematics, then χ^2/DoF is drawn from a χ^2 distribution with mean 1 and variance 2/DoF [2].
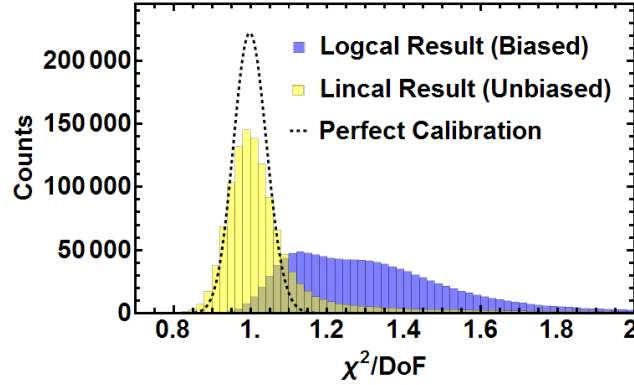
Figure 6-11: Histograms of the distributions of χ^2/DoF of logcal results (mean 1.31) and lincal results (mean 1.05, median 1.01), together with the theoretical distribution of χ^2/DoF (mean 1). They contain one night of data in a 12.5MHz frequency band (21:00-5:00 in Figure 6-10). We evaluate χ^2/DoF for every 5.3 seconds and every 49 kHz. We set the flagging threshold to $\chi^2 = 1.2$, and 80% of the lincal result is below the threshold (majority of the 20% flagged data have χ^2 much larger than 2, thus not shown in this figure). Among the data that is not flagged, 85% is accounted for by the theoretical χ^2 distribution. The right tail in lincal's distribution is due to the noise model sometimes underestimating the noise in order to minimize false negatives in the flagging process. The fact that χ^2/DoF for lincal is so close to the theoretical distribution means that both the instrument and the calibration algorithms are working exactly as we expect.

With a smooth model for σ_{ij} which we describe below, we compute χ^2/DoF for the results of rough calibration, log calibration, and linear calibration using all of our data. The χ^2 distributions of our calibrations for one day's data are shown in Figures 6-10 and 6-11. Each calibration algorithm significantly reduces the χ^2/DoF, and lincal's produces a distribution of χ^2/DoF consistently centered around 1. We automatically flag any data with χ^2/DoF larger than 1.2, which accounts for about 20% of the data. Among the data that is not flagged, 85% is accounted for by the theoretical χ^2 distribution. The 15% in the right tail is mostly attributable to a slightly optimistic noise model designed to avoid underestimating χ^2. This close agreement between predicted and observed χ^2-distributions for the lincal results suggests that except during periods that get automatically flagged, our instrument and analysis pipeline is free from significant systematic errors. The fully automatic nature of our calibration pipeline and data quality assessment is encouraging for future instruments with data volume too large for direct human intervention.

Calculating χ^2/DoF for flagging and data quality assessment requires an accurate model of noise in the measured visibilities. To compute the noise σ_{ij}, we approximate σ_{ij}^2 by $\langle \sigma^2 \rangle$, where the average is over all baselines. This assumption that all antennas have the same noise properties drastically deceases the computational cost of calculating χ^2/DoF. Because we have 10^3 baselines, and the variation of σ_{ij} between baselines is less than 20% (due to slightly different amplifier gains), this approximation should cause only about a 1% error in the final χ^2/DoF.

To compute $\langle \sigma^2 \rangle$, we perform linear regression on each visibility v_{ij} over one minute to obtain its estimated variance σ_{ij}^2, and then average all σ_{ij}^2 to obtain σ^2. Thus we have $\langle \sigma^2 \rangle$ at all frequencies every minute. Before we plug $\langle \sigma^2 \rangle$ into Equation 6.6, we model it as a smooth and separable function: $\langle \sigma^2 \rangle(f,t) = F(f)T(t)$, where $F(f)$ and $T(t)$ are polynomials. The smooth model has three advantages. The first is that it is physically motivated to model thermal fluctuation as a smooth and separable function. Secondly, a smooth noise model makes the χ^2/DoF a much more sensitive flagging device. Theoretically, χ^2/DoF should not rise above 1 when unwanted radio events such as radio frequency interference (RFI) occur, because they are far field

signals that do not violate any redundancy. However, since RFI events make both the signal and noise stronger, by demanding a smooth noise model, the $\langle \sigma^2 \rangle$ we use will underestimate the noise during RFI events and give abnormally high χ^2/DoF, which can then be successfully flagged with the $\chi^2/\text{DoF} < 1.2$ threshold. Thirdly, seasonal changes aside, the noise model is expected to largely repeat itself from day to day, so for future experiments that will operate for years, it suffices to use the same model repeatedly without recomputing σ_{ij} *in situ* for all the data. Thus, by using a smooth noise model, one can drastically reduce the occurrence of false negatives (since it is better to flag good data than it is to fail to flag bad data) as well as the computational cost of calculating χ^2/DoF.

6.3.1.5 Optimal Filtering of Calibration Parameters

While the above-mentioned estimates of the calibration parameters that we obtain from redundant baseline calibration vary over time and frequency, much of that variation is due to the noise in raw data. To minimize the effect of instrumental noise on the calibration parameters, we would like to optimally average information from nearby times and frequencies to estimate the calibration parameters for any particular measurement.

As we will show below, the optimal method for performing this averaging is Wiener filtering. In the rest of this section, we first measure the power spectrum of the calibration parameters over time and frequency, and make a determination of how to decompose this into contributions from signal (true calibration changes) and noise. We then weight the Fourier components in a way that is informed by their signal-to-noise ratio, and quantify how this Wiener filtering procedure improves upon more naive averaging over time and/or frequency. Finally, we discuss the implications for how regularly (in time and frequency) we should calibrate. It is important to note that while these methods are applied only to MITEoR below, they are applicable to any current or future experiment.

We model the measured calibration parameter $g_i(f, t)$ for the ith antenna as the sum of a true calibration parameter $s_i(f, t)$ (the "signal") and uncorrelated noise

$n_i(f, t)$:

$$g_i(f, t) = s_i(f, t) + n_i(f, t). \tag{6.8}$$

We choose our estimator $\hat{g}_i(f, t)$ of the true calibration parameter $s_i(f, t)$ to be a linear combination of the observed calibration parameters g_i at different times and frequencies:

$$\hat{g}_i(f, t) \equiv \int \int W(f, t, f', t') g_i(f', t') df' dt' \tag{6.9}$$

for some weight function W. We optimize the estimator $\hat{g}_i$ by choosing the weight function W that minimizes the mean-squared estimation error $\langle |\hat{g}_i(f, t) - s_i(f, t)|^2 \rangle$. Assuming that the statistical properties of the signal and noise fluctuations are stationary over time[18], all correlation functions become diagonal in Fourier space:

$$
\begin{aligned}
\langle \tilde{s}_i(\tau, \nu)^* \tilde{s}_i(\tau', \nu') \rangle &= (2\pi)^2 \delta(\tau' - \tau) \delta(\nu' - \nu) S(\tau, \nu), \\
\langle \tilde{n}_i(\tau, \nu)^* \tilde{n}_i(\tau', \nu') \rangle &= (2\pi)^2 \delta(\tau' - \tau) \delta(\nu' - \nu) N(\tau, \nu), \\
\langle \tilde{s}_i(\tau, \nu)^* \tilde{n}_i(\tau', \nu') \rangle &= 0,
\end{aligned}
\tag{6.10}
$$

where tildes denote Fourier transforms and S and N are the power spectra of signal and noise, respectively. This means that the optimal filter becomes a simple multiplication $\hat{\tilde{g}} = \tilde{W} \tilde{g}$ in Fourier space, corresponding to the weighting function $\tilde{W}(\tau, \nu)$ that minimizes the mean-squared error

$$\langle |\tilde{W}(\tau, \nu) \tilde{g}_i(\tau, \nu) - \tilde{s}_i(\tau, \nu)|^2 \rangle. \tag{6.11}$$

Requiring the derivative of this with respect to $\tilde{W}$ to vanish gives the Wiener filter [232]

$$\tilde{W}(\tau, \nu) = \frac{S(\tau, \nu)}{S(\tau, \nu) + N(\tau, \nu)}. \tag{6.12}$$

Since S and N are to a reasonable approximation independent of the antenna number i, we have dropped all subscripts i for simplicity. Back in real space, the optimal

---

[18]We perform this analysis on data over 12 MHz and two hours, where the signal and noise power are empirically found to be approximately time-independent.

estimator $\hat{g}_i$ for the ith calibration parameter is thus g_i convolved with the 2D inverse Fourier transform of $\tilde{W}$.

To demonstrate this technique, we show the above process carried out in the time dimension in Figure 6-12. In practice we perform the analysis on time and frequency dimensions simultaneously through a 2D FFT. The noise power spectrum $N(\nu)$ is seen to be constant to an excellent approximation, corresponding to white noise (uncorrelated noise in each sample). The signal power spectrum $S(\nu)$ is seen to be well fit by a combination of two power laws: $S(\nu) \approx (\nu/2.9 \times 10^{-5}\,\text{Hz})^{-2.7} + (\nu/4.8 \times 10^{-17}\,\text{Hz})^{-0.46}$. The optimal convolution kernel W is seen to perform a weighted average of the data on the timescale of roughly $200\,\text{s}$ and frequency scale of $0.15\,\text{MHz}$, giving the greatest weight to nearby times and frequencies, resulting in an order-of-magnitude noise reduction.

To quantify the effectiveness of the obtained filter compared to naive "boxcar" averages, we use the 2D power spectrum and noise floor of the calibration parameters obtained from real data to simulate many realizations of calibration parameters $g(f,t) = s(f,t) + n(f,t)$, apply various averaging/convolution schemes $W(f,t)$ on the simulated data, and compare their effectiveness by computing the RMS error $\langle |(W \star g)(f,t) - s(f,t)|^2 \rangle$ normalized by $\langle |n|^2 \rangle$. Due to our limited frequency bandwidth as well as frequent RFI contamination, power spectrum modeling in the frequency dimension is very challenging, so the frequency Wiener filter appears to be less effective than the time filter. In Table 6.2 we list the normalized noise powers using frequency Wiener filter, time Wiener filter, 2D Wiener filter, as well as traditional boxcar averaging, and the 2D Wiener filter produces results three times less noisy than that of the traditional boxcar averaging.

We have described how to optimally average calibration parameters when we calibrate very regularly in time and frequency. For a future instrument such as an omniscope with 10^6 antennas, calibration will pose a serious computational challenge, so it is important to know what the minimal frequency one needs to calibrate the instrument. The above analysis conveniently provides an answer to this question. As
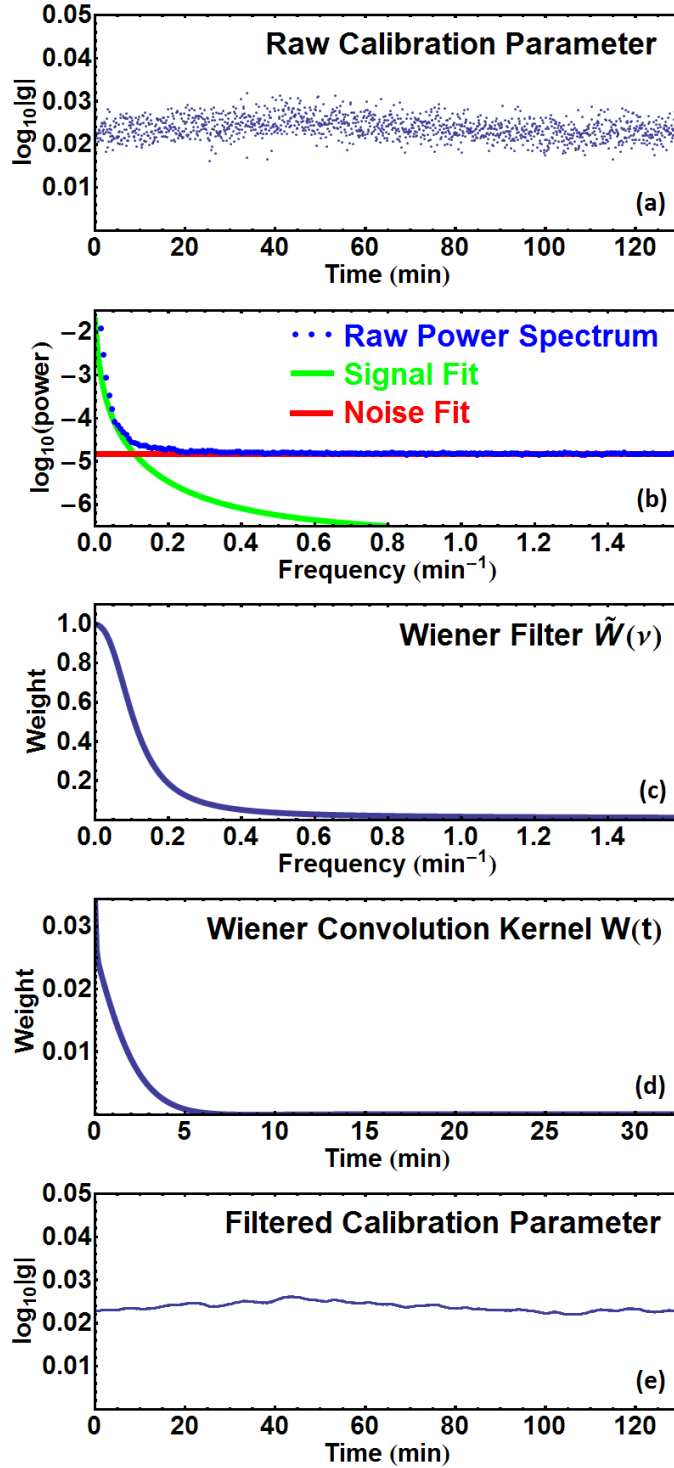
Figure 6-12: Illustration of 1D Wiener filtering of calibration parameters at different times. Panel (a) shows the amplitude of calibration parameters measured for one antenna over two hours. Panel (b) shows that the average power spectrum across all antennas (blue dots) is well fit by a white noise floor (red horizontal line) plus a sum of two power laws (green curve). Panel (c) shows the Wiener filter in frequency domain computed using Eq. 6.3.1.5 and the power spectra from panel (b). Panel (d) shows the Wiener convolution kernel in the time domain, the Fourier transform of the filter in Panel (c). Panel (e) shows the best estimates of the true calibration amplitude. The effectiveness of this filter is compared with that of other filters in Table 6.2.

Averaging method	Relative noise power
No average	1
Frequency Wiener filter	0.33
Time Wiener filter	0.12
Time and frequency Wiener filter	0.09
Time and frequency boxcar average	0.32

Table 6.2: Wiener filtering reduces the noise contribution to the calibration parameters by an order of magnitude. This table lists residual noise power (normalized by original noise power) after applying various filters to average the amplitude calibration parameters in time and/or frequency. The optimal two-dimensional Wiener filter indeed performed the best, lowering the noise power by an order of magnitude. In comparison, the naive boxcar average, using the characteristic scales of the optimal Wiener filter (200 s and 0.15 MHz), has more than 3 times residual noise power than the Wiener filtered result.

shown in the second panel of Figure 6-12, the signal[19] is band limited. By the Nyquist theorem, one needs to sample with at least double the frequency of signal bandwidth, so in our case we could measure the calibration parameters without aliasing problems as long as we calibrate once per minute. Calibrating more frequently than this simply helps average down the noise. Although this one-minute timescale depends on the temporal stability characteristics of the amplifiers and other components used in our particular experiment, it provides a useful lower bound on what to expect from future experiments whose analog chains are even more stable.

6.3.2   Absolute Calibration

The absolute calibration of the array involves two separate tasks. One is to find the overall gain and to break phase degeneracies that redundant baseline calibration is unable to resolve, and the other to calibrate fixed properties of the instrument such as the orientation of the array and shape of the primary beam. The former is done by comparing the data to a sky model comprised of the global sky model (GSM) [54] and published astronomical catalogs (see Jacobs et al. 104 for example). The latter is done using bright point sources with known positions. While we can take advantage of the

---

[19]We only show results for amplitude calibration parameters for brevity, as the phase calibration results have nearly identical power spectrum.

extremely high signal-to-noise data in the ORBCOMM channels (around 137 MHz), thanks to the dynamic range provided by our 8 bit correlator, it is important to note that all the algorithms described here are applicable to astronomical point sources as well.

This section is divided into three parts. The first part describes how we use prior knowledge of the sky to break the degeneracies in redundant calibration results, a vital step to obtain usable data products. The second and third part each describe one aspect of absolute calibration using satellite data: primary beam measurement and array orientation.

6.3.2.1   Breaking Degeneracies in Redundant Calibration

Redundant calibration alone cannot produce directly usable data products, due to the degeneracies intrinsic to the algorithms. There is one degeneracy in the amplitude of the calibration coefficients g_i, since scaling the amplitude of everything up by a common factor does not violate any redundancy (the degeneracies discussed here are per frequency and per time, as are the calibration solutions). There are three degeneracies in phase, corresponding to three degrees of freedom in a two dimensional linear field (see Appendix 6.A for a detailed discussion). In general, breaking these degeneracies requires prior knowledge of the sky. In this section, we briefly describe our algorithm that uses the global sky model (GSM) of de Oliveira-Costa et al. [54] to remove these degeneracies. Doing so requires efficiently simulating the response of the instrument to the GSM; we summarize a fast algorithm for doing so in Appendix 6.C. We defer detailed comparison of our data and the GSM to a future publication.

Our degeneracy removal procedure is an iterative loop that repeats two steps. The first step is to fit for the amplitude degeneracy factor. The knowledge of the GSM and bright point sources give us a set of model visibilities, m_{ij}^a, where index a denotes different modeled components such as the GSM or Cygnus A. A linear combination of these models should be able to fit our measurements[20]. Thus we fit for the weights

---

[20]We allow each model to have a separate weight to guard against potential calibration offsets between existing models.

w_a of the models by minimizing

$$\left| v_{ij} - \sum_a w_a m_{ij}^a \right|^2 . \tag{6.13}$$

The second step is to break the degeneracy in redundant phase calibration by fitting for the degeneracy vector $\boldsymbol{\Phi}$ and the constant ψ defined in Appendix 6.A. We assume that the error in the first step's fitting is mostly due to the phase degeneracies, so we take the w_a from step one and fit for $\boldsymbol{\Phi}$ and ψ by minimizing

$$\left| \arg(v_{ij}) - \arg \left(\sum_a w_a m_{ij}^a \right) - \boldsymbol{d}_{i-j} \cdot \boldsymbol{\Phi} - \psi \right|^2 , \tag{6.14}$$

where $\boldsymbol{d}_{i-j}$ is the position vector for baseline $i - j$.

Note that the two fitting processes described above are not independent of one another, so we repeat these steps until convergence is reached. We find that in practice, the errors converge within two iterations. Our preliminary result is illustrated in Figure 6-13, which shows that the data agree very well with current models.

6.3.2.2 Beam Measurement Using ORBCOMM Satellites

In general, *in situ* measurements of antenna primary beams over large fields of view pose a challenge to 21 cm cosmology, as primary beam uncertainties are intimately related to calibration, imaging, and catalog flux uncertainties [103]. Motivated by these difficulties, Pober et al. [181] present a solution that uses celestial point sources and assumes reflection symmetry of the beam, whereas Neben, Bradley, and Hewitt (in preparation) demonstrate high dynamic range beam measurement using the constellation of ORBCOMM satellites. Here, we present *in situ* primary beam measurements of the MWA bow-tie antennas using the ORBCOMM constellation. We take advantage of both the high signal-to-noise ratio of ORBCOMM signals, and of our full cross-correlation measurements (rather than auto-correlations alone) to determine the beam.

In order to measure their primary beam profile $B_{\mathrm{mwa}}(\hat{\mathbf{r}})$, we compare measure-
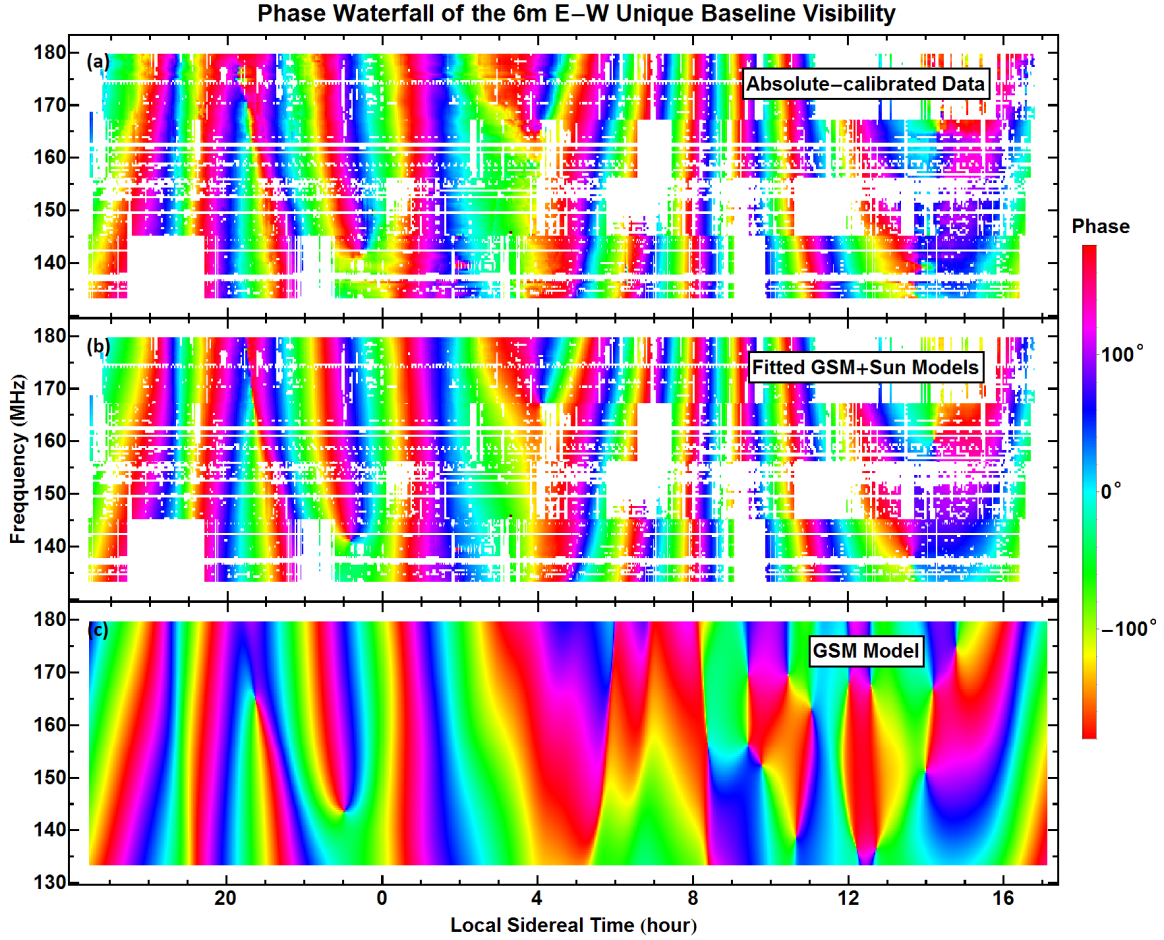
Figure 6-13: Waterfall plots of phases on the 6 m E-W baseline. These show that our absolute calibration successfully matches the data (panel (a)) with a linear combination of global sky model and known point sources, including the Sun (panel (b)). Panel (c) shows the global sky model alone. The white areas are flagged out using χ^2 criterion described in Figure 6-10. Each plot is stitched together using four independently measured and calibrated frequency bands, aligning local sidereal time. Thus the discontinuities between hours 4 and 12 are due to the Sun rising at different local sidereal times on different days of our observing expedition.

ments with MWA antennas to simultaneous measurements with simple center-fed dipoles, whose beam pattern B_{dipole} is known analytically. When there is a single extremely bright point source in the sky, such as an ORBCOMM satellite, we can compute the ratio of the visibilities of select baselines to obtain the ratio of the MWA antenna beam to the analytically known dipole antenna beam — thus determining the MWA antenna beam itself. To perform this analysis, two dipole antennas, one orientated along the x-polarization axis of the array and the other along the y-polarization axis, are added to the array and cross-correlated with all other MWA antennas.

The rationale behind this technique is as follows. At an angular frequency ω, the electric field from a sky signal at the position of a receiving antenna can be encoded in the Jones vector $\boldsymbol{S}(\hat{\boldsymbol{k}})$, where $\hat{\boldsymbol{k}}$ is the position vector of the source [46]. With a primary beam matrix $\mathbf{B}_j(\hat{\boldsymbol{k}})$, the signal measured by the jth antenna at position $\boldsymbol{r}_j$ is

$$s_j = \int e^{-i[\boldsymbol{k}\cdot\boldsymbol{r}_j+\omega t]}\mathbf{B}_j(\hat{\boldsymbol{k}})\boldsymbol{S}(\hat{\boldsymbol{k}})\,\mathrm{d}\Omega. \tag{6.15}$$

When a single ORBCOMM satellite is above the horizon[21], its signal strength is so dominant at its transmit frequency that $\boldsymbol{S}(\hat{\boldsymbol{k}})$ becomes well-approximated by a point source at the satellite's location. The measured signal can then be written as:

$$s_j \approx e^{-i[\boldsymbol{k}_s\cdot\boldsymbol{r}_j+\omega t]}\mathbf{B}_j(\hat{\boldsymbol{k}}_s)\boldsymbol{S}_s, \tag{6.16}$$

where $\boldsymbol{k}_s$ is the wave vector of the satellite signal, and $\boldsymbol{S}_s$ is the Jones vector encoding the satellite signal strength.

If we limit our attention to either x-polarization or y-polarization and approximate the off diagonal terms of $\mathbf{B}(\hat{\boldsymbol{k}})$ as zero, the visibility for two antennas can be written as

$$v_{jk} \approx S^2 B_j(\hat{\boldsymbol{k}}_s)^* B_k(\hat{\boldsymbol{k}}_s)e^{-i\boldsymbol{k}_s\cdot(\boldsymbol{r}_k-\boldsymbol{r}_j)}. \tag{6.17}$$

If we take one visibility v_{ij} formed by correlating a simple center-fed dipole with an MWA bow-tie antenna and another visibility v_{kl} for the same baseline vector formed

---

[21]There is typically more than one ORBCOMM satellite above the horizon at any one time, but they are coordinated so that they do not transmit in the same frequency band simultaneously.

by correlating two MWA antennas, then their ratio is simply

$$\frac{|v_{ij}|}{|v_{kl}|} \approx \frac{|B_{\mathrm{mwa}}(\hat{\boldsymbol{k}}_s)|}{|B_{\mathrm{dipole}}(\hat{\boldsymbol{k}}_s)|}, \tag{6.18}$$

because the satellite intensity S and one MWA beam factor B_{mwa} all cancel out, and the phase factor $e^{-i\boldsymbol{k}_s \cdot (\boldsymbol{r}_j - \boldsymbol{r}_i)}$ is removed due to taking absolute values of the visibilities. This means that when a single point source dominates the sky, the ratio of visibility amplitudes is simply the ratio of the antenna beams at the direction of the point source. Since we already know the beam B_{dipole} of a center-fed dipole over a ground screen, we can directly infer the magnitude of MWA primary beam $|B_{\mathrm{mwa}}(\hat{\boldsymbol{k}}_s)|$.

In order to fully map out the MWA primary beam, we need to take data during many satellite passes until we have direction vectors that densely cover the entire sky. Satellite signals from 27 ORBCOMM satellites at 5 frequencies in the range of 137.2-137.8 MHz were identified. Their orbital elements are publicly available[22], so we can calculate $\hat{\boldsymbol{k}}_s(t)$ straightforwardly. With 40 hours of data taken at the frequencies of interest, we were able to obtain 248 satellite passes, shown in Figure 6-14.

We compared our measurements of the MWA primary beam using Equation 6.18 to numerical calculations using the FEKO electromagnetic modeling software package. Fixing an azimuth angle ϕ, we can plot and compare the simulated and measured beam at different polar angles θ (the angle between the direction vector and zenith). Figure 6-15 shows how the beam changes with θ for four different ϕ-values, where $\phi = 0$ correspond to North and increases clockwise. Our measurements of the MWA beams are seen to agree well with the numerical predictions for both polarizations. The small differences between the predicted and measured beams are larger than the statistical noise, implying that the main limitation is not noise but one or more of the above-mentioned approximations, or approximations in the electromagnetic antenna modeling.

---

[22] We obtained the TLE files from CelesTrak, a company that archives TLEs of many civil satellites.
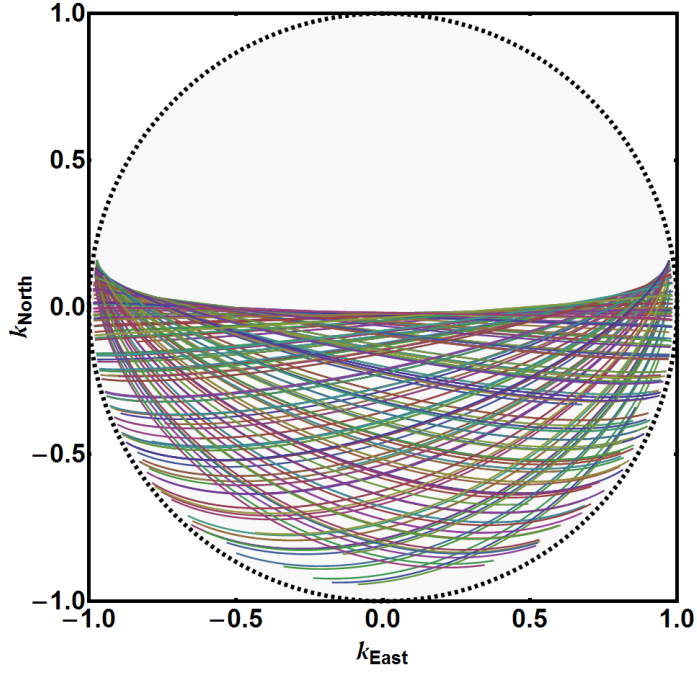
Figure 6-14: Projected trajectories of 248 passes of ORBCOMM satellites over 40 hours. With these passes we obtain sufficiently dense sampling of the MWA antennas primary beam that we can robustly map its response, especially at high elevations where the response is strongest. With a map of the southern half of the primary beam, we can use the reflection symmetry of the antennas to infer the entire beam at the ORBCOMM transmission frequencies. Each curve is a satellite pass projected onto the x-y plane, and the different colors specify sets of data taken at different times.
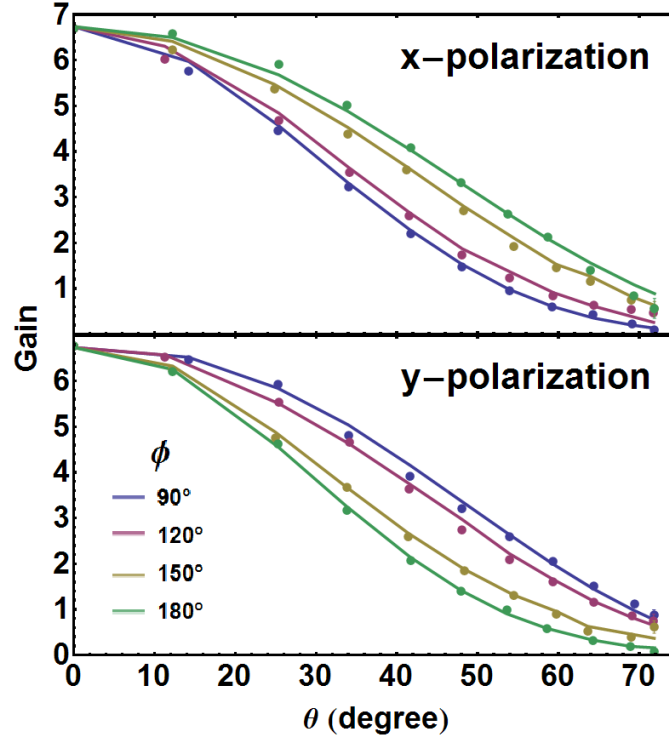
Figure 6-15: Measured MWA primary beam patterns compared to those obtained from numerical modeling. The two panels show the predictions (curves) and measurements (points) of the primary beam for the x-polarized and y-polarized MWA antennas. Each curve shows how the primary beam changes with the polar angle θ for a fixed azimuth angle ϕ. To reduce noise, the measurements have been averaged in 10 square degree bins.

6.3.2.3 Calibrating Array Orientation Using ORBCOMM Satellites and the Sun

The orientation of the array is very important, because the degeneracy removal process relies on the predicted measurement for each unique baseline, which in turn relies on precise knowledge of the baselines' orientations. Although we measured the relative position of each antenna to millimeter level precision with a laser-ranging total station, we did not measure the absolute orientation of the array to better than the $\sim 1°$ accuracy obtainable with a handheld compass. To improve upon this crude measurement, we make use of both the known positions of both ORBCOMM satellites and the Sun. As we show in Figure 6-16, the exceptional signal-to-noise in the ORBCOMM data allows us to fit for a small array rotation as a first order correction to a model based on our crude measurement. Our method for finding the true orientation of the array is as follows. For a given baseline during the peak few minutes of an ORBCOMM satellite pass at frequency ν, we measure a phase $\phi(t)$. We also know the satellite's position vector $\boldsymbol{k}(t)$. However, we only have crude knowledge of baseline vector $\boldsymbol{d}_0$ in units of wavelength, where vectors are in horizontal coordinates with x, y, z that correspond to south, east and up. We can therefore only compute a crude prediction of the phase measurement

$$\phi_0(t) = 2\pi \boldsymbol{k}(t) \cdot \boldsymbol{d}_0. \tag{6.19}$$

We assume that the difference between the measurement $\phi(t)$ and our crude prediction $\phi_0(t)$ is due to a small angle rotation of the baseline vector $\boldsymbol{d}_0$ around the axis $\boldsymbol{\theta} = (\theta_x, \theta_y, \theta_z)$ by an angle $\theta = |\boldsymbol{\theta}|$, ignoring a constant cable length delay.[23] In the small

---

[23]Here it is important to use data before redundant baseline calibration to avoid phase degeneracy. We remove the phase delay from cables by allowing a constant offset that matches $\phi(t_M)$ with the crude prediction at time t_M when the satellite has the strongest signal during the pass.
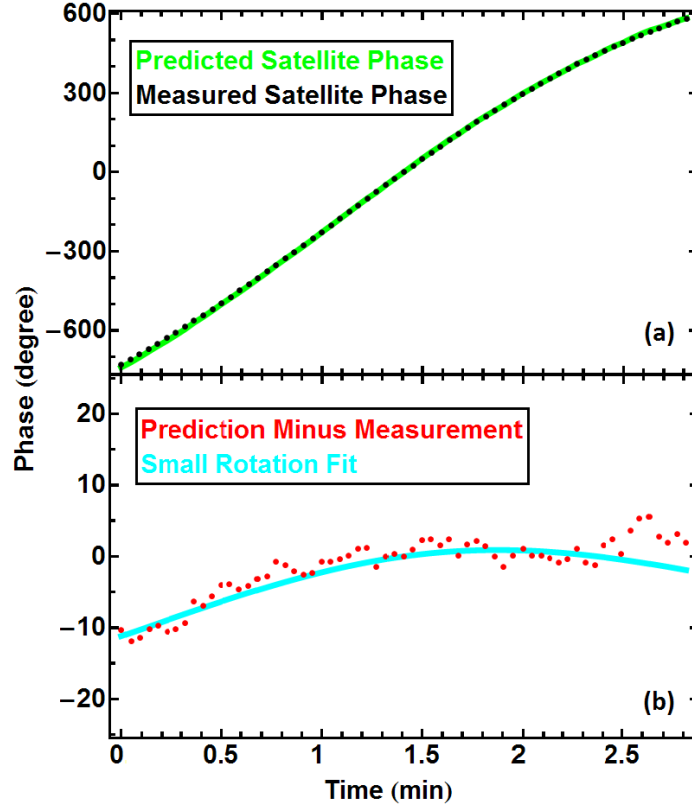
Figure 6-16: Illustration of using the high signal-to-noise ORBCOMM data to calculate any small rotation in the array relative to the field-measured orientation. Panel (a) shows the rapidly wrapping phase of the raw data (black) from one baseline at the ORBCOMM frequency during the peak three minutes of a single satellite pass. In green, we see the predicted values computed with the field-measured array orientation and publicly available satellite positions. The residual between the model and the data is plotted in red points in panel (b). Finally, the cyan curve shows the best fit using small angle rotations of the array. In practice we use hundreds of satellite passes and all the baselines to obtain a single accurate fit for the true orientation of the array.

θ regime, we have that

$$\phi(t) - \phi_0(t) = 2\pi \boldsymbol{k}(t) \cdot (\mathbf{R}(\boldsymbol{\theta}) \cdot \boldsymbol{d}_0 - \boldsymbol{d}_0)$$

$$\approx 2\pi \boldsymbol{k}(t) \cdot (\boldsymbol{\theta} \times \boldsymbol{d}_0)$$

$$= 2\pi (\boldsymbol{d}_0 \times \boldsymbol{k}(t)) \cdot \boldsymbol{\theta}, \qquad (6.20)$$

where $\mathbf{R}(\boldsymbol{\theta})$ is the rotation matrix. Since we have a set of equations each representing a different time, the problem of finding $\boldsymbol{\theta}$ can be reduced to that of finding a least squares fit. With 117 satellite passes, we obtained the following best fit for the array rotation around the vertical axis:

$$\theta_z^{\mathrm{sat}} = 0.66° \pm 0.0005°_{\mathrm{stat}} \pm 0.07°_{\mathrm{sys}}.$$

While this method is very precise for solving the main problem we were worried about—the direction of North (θ_z) which we approximated in the field with a hand-held compass—it is less useful for measuring rotations of the array in the other two directions. Our instrument's absolute timing precision is only ~ 0.5 seconds, which makes it hard to distinguish rotations about the North-South axis from timing errors, as most ORBCOMM passes are East-West. This issue can of course be easily addressed in future experiments; for our experiment, we solve it using a more slowly moving bright point source: the Sun.

By using one day of solar data at $139.3\,\mathrm{MHz}$, we obtained

$$(\theta_x, \theta_y, \theta_z)^\odot = (-0.08°, -0.12°, 0.672°)$$

$$\pm (0.01°, 0.03°, 0.004°)_{\mathrm{stat}}$$

$$\pm (0.04°, 0.003°, 0.005°)_{\mathrm{sys}}.$$

Although solar data is noisier, in part because the Sun is not as bright as the ORBCOMM satellites in a given channel, timing errors are no longer important. These results agree with and complement the satellite-based results and allow us to con-

fidently pin down the orientation of the array and thus improve the quality of the calibration of all of our data. The excellent agreement between the independent measurements $\theta_z^{\mathrm{sat}} \approx 0.66°$ and $\theta_z^{\odot} \approx 0.67°$ provides encouraging validation of both the satellite and solar calibration techniques.

6.3.3 Systematics

As we discussed in section 6.3.1.4, most of our data are well-calibrated with $\chi^2/\mathrm{DoF} <$ 1.2, which means that any systematic effects should lie well below the level of the thermal noise. In this section we aim to identify all the systematic effects present in the system, and describe our efforts to quantify and, whenever possible, remove them. The systematics can be categorized into two groups:

1. Signal-dependent systematics that grow as the signal becomes stronger, such as cross-talk, antenna position errors and antenna orientation errors.

2. Signal-independent systematics, such as radio frequency interference (RFI) from outside or inside the instrument.

Below we find a strict upper bound of 0.15% for the signal-dependent component, as well as a signal-independent component which is easy to remove.

To quantify signal-dependent systematics, we again use ORBCOMM satellite data. Because the ORBCOMM signals are 10^3 times brighter than astronomical signals, and we know that any signal-independent systematics must be weaker than the astronomical signals (otherwise they would have been blatantly apparent in the data), any signal-independent systematics must be negligible compared to the ORBCOMM signal. We therefore investigate how the discrepancies between calibrated visibilities and the models for each unique baseline depend on ORBCOMM signal strength. We define the average fitting error per baseline at a given time and frequency to be

$$\epsilon = \langle |v_{ij} - y_{i-j} g_i^* g_j| \rangle, \tag{6.21}$$

which is a combination of antenna noise and systematic errors. If we compute ϵ at different times with different signal strength and compute its signal dependence, we can derive an upper bound on the signal dependence of systematic errors. To do this, we take all data at the ORBCOMM satellite frequency over a day and compute ϵ after performing redundant calibration. We then bin the ϵ-values according to the average signal strength $s = \langle |y_{i-j}| \rangle$, and obtain the results shown in Figure 6-17[24]. The result is seen to be well fit by a constant noise floor plus a straight line $\epsilon \approx 0.0015s$. This slope implies that the combined effect of all signal-dependent systematic effects is at most 0.15%. This is merely an upper bound on the systematics, since it is possible that the increase in ϵ is mainly due not to systematics but to an increase in instrumental noise caused by an increase in the system temperature during the ORBCOMM passes.

There is one signal-dependent systematic that is not included in the above analysis: deviation from redundancy caused by imperfect positioning of antenna elements. This is because the data we used to derive the upper bound is always dominated by a single point source, the ORBCOMM satellite, and redundant calibration cannot detect any deviation of antenna position when the sky is dominated by a single point source[25]. We have two ways to quantify the error in our data due to antenna position errors. Firstly, the laser-ranging measurements of antenna positions in the field indicate an average of $0.037 \, \text{m}$ deviation from perfect redundancy, which translates to about 2% average error in phase on each visibility. Since the deviations are in random directions, the variance of phase error in the unique baseline fits should be brought down by a factor equal to the number of redundant baselines, resulting in phase errors much less than 1% for most of the unique baselines. Secondly, although satellite calibration cannot detect position error in a given snapshot, over time the position errors would

---

[24]Another way of describing these data points is that, if we look at the third panel in Figure 6-9, we are plotting the average small spread in each unique baseline group versus the radius of the circle, and as the satellites pass over, both the circle size and the amount of average spread change over time, forming the data set in question.

[25]This is because for any arbitrary position deviation $\Delta \boldsymbol{r}_i$ for antenna i, one can add a phase equal to $\boldsymbol{k} \cdot \Delta \boldsymbol{r}_i$ to the calibration parameter g_i to perfectly "mask" this deviation. Note that this "mask phase" depends on $\boldsymbol{k}$ and thus changes rapidly over time when the ORBCOMM satellite moves across the sky.
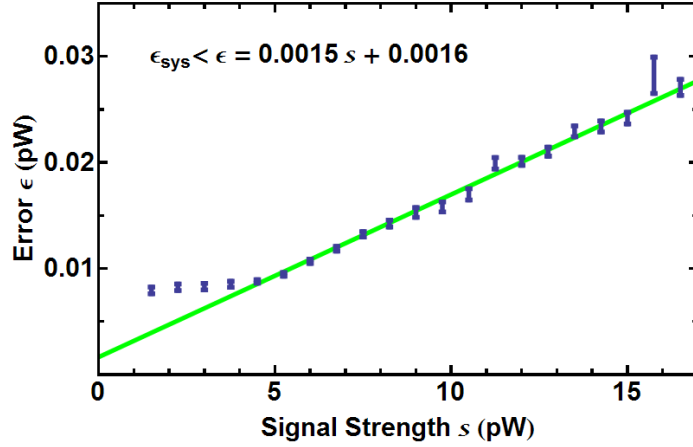
Figure 6-17: Signal-dependent systematic error and its linear fit. By comparing the modeled and calibrated visibilities during ORBCOMM satellite passes, we conclude that signal-dependent systematic effects account for no more than 0.15% of our measurement. We calculate the average fitting error per baseline $\epsilon = \langle |v_{ij} - y_{i-j} g_i^* g_j| \rangle$ and the average signal strength $s = \langle |y_{i-j}| \rangle$ binned over one day's data (blue points). The green line fits the data points above the noise floor. While many systematic errors, such as cross-talk, can contribute to the fitting error in addition to thermal noise, the best-fit slope of 0.0015 puts an upper bound on the sum of all signal-dependent errors. Since the ORBCOMM signal is so strong, any signal-independent systematic errors are negligible in this analysis. The high noise floor of ~ 0.01 pW is due to our digital tuning in the ORBCOMM frequency channels to maximize dynamic range.

create very rapidly changing calibration parameters, which we do not observe in our data. Lastly, a formalism exists [124] to treat errors in antenna placement as small perturbations when redundantly calibrating, although though we did not need to take advantage of this technique for the present paper.

We first identified a signal-independent systematic when we obtained consistent $\chi^2/\mathrm{DoF} \sim 4$ for much of our data[26], which means that the fitting error was on average twice as large as the thermal noise in each visibility. This implies a systematic (or a combination of systematics) at the level of $10^{-6}\,\mathrm{pW/kHz}$, about 10% of the total astronomical signal. Given the above analysis, we can exclude the possibility of any signal-dependent explanations such as cross-talk between channels or antennas. While we are unable to offer any conclusive explanation of this systematic, it appears consistent with persistent near-field RFI, perhaps originating from our electronics. Fortunately, we found this additive signal to vary only very slowly over time, typically remaining roughly constant over 5-minute periods, which made it easy to remove. After calibrating the data with logcal, we average the fitting errors $\epsilon_{ij} = \langle v_{ij} - y_{i-j}g_i^* g_j \rangle_t$ over time and subtract them from the data before we run logcal again. We perform the averaging over 5 minute segments, corresponding to 112 independent time samples, and iterate the calibration-subtraction process three times. This corresponds to less than a 1% increase in the number of effective calibration parameters we fit for during logcal. Because many baselines probe the same unique baseline, the procedure described above exploits the redundancy of the array to robustly remove this slowly varying, signal-independent systematic, leaving us with $\chi^2/\mathrm{DoF} \sim 1$.

6.4 Summary and Outlook

We have described the MITEoR experiment, a pathfinder "omniscope" radio interferometer with 64 dual-polarization antennas in a highly redundant configuration. We have demonstrated a real-time precision calibration pipeline with automatic data

---

[26]This was before we obtained a consistent $\chi^2/\mathrm{DoF} \sim 1$ in Section 6.3.1.4, which occurred after we were able to remove the systematic described here.

quality monitoring that uses χ^2/DoF as a data quality metric to ensure that redundant baselines are truly seeing the same sky. We have also implemented various instrumental calibration techniques that utilize the ORBCOMM constellation of satellites to measure the primary beam and precise orientation of the array. Our success bodes well for future attempts to perform such calibration in real-time instead of in post-processing, and thus clears the way for FFT correlation that will make interferometers with $\gtrsim 10^3$ antennas cost-efficient by reducing the computational cost of correlating N antennas from an N^2 scaling to an $N \log N$ scaling. It also suggests that the extreme calibration precision required to reap the full potential of 21 cm cosmology is within reach.

The various calibration techniques that MITEoR successfully demonstrates are now being incorporated into the much more ambitious HERA project[27] [184], a broad-based collaboration among US radio astronomers from the PAPER, MWA, and MITEoR experiments. Our results are also pertinent to the design of the SKA low-frequency aperture array[28]. HERA plans to deploy around 331 14-meter dishes in a close-packed hexagonal array in South Africa, giving a collecting area of more than 0.05 km^2, virtually guaranteeing not only a solid detection of the elusive cosmological 21 signal but also interesting new clues about our cosmos.

6.A   Appendix: Phase Degeneracy in Redundant Calibration

Both of our redundant baseline calibration algorithms, logcal and lincal (see Section 6.3.1.3), have the same set of phase degeneracies that require additional absolute calibration that must incorporate knowledge of the sky. When calibrating a given

---

[27]http://reionization.org/
[28]http://skatelescope.org/

unique baseline, the quantity that logcal minimizes is

$$\sum_{jk} |(\theta_{j-k} - \phi_j + \phi_k) - \arg(v_{jk})|^2, \tag{6.22}$$

where we define $\theta_{j-k} \equiv \arg(y_{j-k}), \phi_j \equiv \arg(g_j)$. Similarly, lincal minimizes

$$\sum_{jk} |(y_{j-k}g_j^*g_k) - v_{jk}|^2$$
$$= \sum_{jk} \left||y_{j-k}g_j^*g_k| \exp\left[i(\theta_{j-k} - \phi_j + \phi_k)\right] - v_{jk}\right|^2. \tag{6.23}$$

Unfortunately, for all values of θ_{j-k} and ϕ_k, one can add any linear field $\boldsymbol{\Phi}\cdot\boldsymbol{r}_j + \psi$ to the ϕ_j across the entire array while subtracting $\boldsymbol{\Phi}\cdot\boldsymbol{d}_j$ from the θ_{j-k} without changing the minimized quantities:

$$\theta'_{j-k} - \phi'_j + \phi'_k \equiv (\theta_{j-k} - \boldsymbol{\Phi}\cdot\boldsymbol{d}_{j-k}) - (\phi_j + \boldsymbol{\Phi}\cdot\boldsymbol{r}_j + \psi)$$
$$+ (\phi_k + \boldsymbol{\Phi}\cdot\boldsymbol{r}_k + \psi)$$
$$= \theta_{j-k} - \phi_j + \phi_k. \tag{6.24}$$

Here $\boldsymbol{r}_j$ is the position vector of antenna j and $\boldsymbol{d}_{j-k} \equiv \boldsymbol{r}_k - \boldsymbol{r}_j$ is the baseline vector for the unique baseline with best-fit visibility y_{j-k}. Thus, the quantities in expressions 6.22 and 6.23 that the calibrations minimize are degenerate with changes to the linear phase field $\boldsymbol{\Phi}$ and the scalar ψ. This means that there are, in general, 4 degenerate phase parameters that need absolute calibration: one overall phase ψ and three related to the three degrees of freedom of the linear function $\boldsymbol{\Phi}$ (which reduces to two for a planar array).

In an ideal instrument, the measured visibilities for a given unique baseline would be

$$y_{i-j} = \int_{k_x,k_y} e^{i\boldsymbol{k}\cdot\boldsymbol{d}_{i-j}} S_B(k_x, k_y) dk_x dk_y, \tag{6.25}$$

where $\boldsymbol{k} = (k_x, k_y, k_z)$ is the wave vector of incoming radiation and $S_B(k_x, k_y)$ is the

338

product of the incoming signal intensity and the primary beam in the direction $\hat{\boldsymbol{k}}$ normalized by kk_z (which comes from the Jacobian of the coordinate transformation from spherical coordinates; see Tegmark and Zaldarriaga 210). When the array is coplanar, we can take an inverse Fourier transform of y_{i-j} and obtain an image of $S_B(k_x, k_y)$. Above we saw that the best fit y_{i-j} computed by logcal and lincal is multiplied by an unknown linearly varying phase $\boldsymbol{\Phi} \cdot \boldsymbol{d}_{i-j}$. Since multiplication in uv space is a convolution in image space, this means that the image generated using these y_{i-j} is the true image convolved with a Dirac delta function centered at $\boldsymbol{\Phi}$, which corresponds to a simple shift by the unknown vector $\boldsymbol{\Phi}$ in the $S_B(k_x, k_y)$ image space.

To calibrate these last few overall phase factors, one can either make sure that bright radio sources line up properly in the image, or match phases between measured visibilities and predicted visibilities, as we described in Section 6.3.2.1. However, there may be another complementary way to remove this phase degeneracy without any reference to the sky. We know that physically the true image $S_B(k_x, k_y)$ is only non-zero within the disk $|k_x^2 + k_y^2|^{1/2} \leq k$ centered around the origin, and a shift caused by $\boldsymbol{\Phi}$ would move this circle off center. This suggests that we should be able to reverse engineer $\boldsymbol{\Phi}$ by looking at how much the image circle has been shifted, without knowing what $S_B(k_x, k_y)$ is. Figure 6-18 demonstrates how the image is shifted by $\boldsymbol{\Phi}$ using simulated data.

Unfortunately, this simple approach to identifying and removing the effect of $\boldsymbol{\Phi}$ suffers from a few complications. By far the most important one is the requirement of very short baselines. In the example in Figure 6-18, the shortest separation between antennas is 0.21λ, and it is easy to show that the sky disk is only clearly demarcated when the shortest separation is less than 0.5λ[29]. This sets a limit on the physical size of each element, which makes achieving a given collecting area proportionately more difficult. As Figure 6-19 shows, the deployed configuration of MITEoR cannot be used to reverse engineer the degeneracy vector $\boldsymbol{\Phi}$ without knowledge of the true

---

[29]This is the 2D imaging counterpart of the well-known fact that, in signal processing, one must sample with a time interval shorter than $0.5\nu^{-1}$ to avoid aliasing in the spectrum of maximum frequency ν.
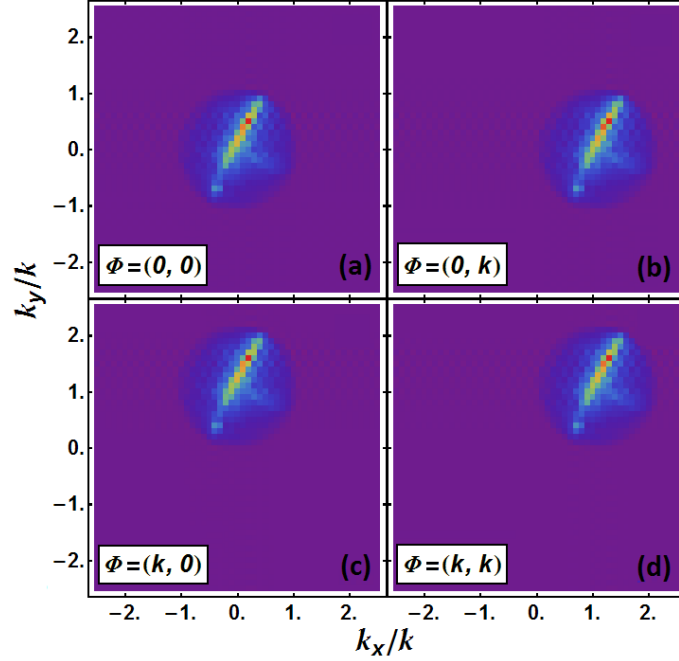
Figure 6-18: Illustration of phase degeneracies shifting the sky image where the sky disk is demarcated. The linear phase degeneracy, which takes the form $\boldsymbol{\Phi} \cdot \boldsymbol{r}_i$ in each antenna for any $\boldsymbol{\Phi}$, corresponds to a shift of the reconstructed image. These simulated images demonstrate shifts of fiducial sky image at 160 MHz caused by four different $\boldsymbol{\Phi}$, where the fiducial array's shortest baseline is 0.2 m. Panel (a) shows the image obtained from visibilities with no $\boldsymbol{\Phi}$, and the sky image is centered at 0. In the other three panels, the sky image is shifted by amount $\boldsymbol{\Phi}$. Even if one has no knowledge of what the true sky is, it is still possible to determine $\boldsymbol{\Phi}$ from where the sky image is centered.
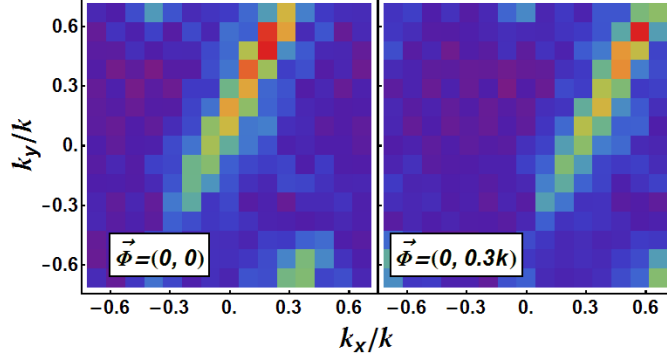
Figure 6-19: Illustration of phase degeneracies shifting the sky image where the sky disk is not demarcated. With any practical array configuration, including that of MITEoR, distinguishing image shifts caused by the $\boldsymbol{\Phi}$-degeneracy becomes significantly more difficult. These images demonstrate shifts of fiducial sky image at 160 MHz just as in Figure 6-18, but with MITEoR's compact configuration where the shortest baseline is 1.5 m. In the left panel, the image is obtained from visibilities with $\boldsymbol{\Phi} = (0, 0)$, and in the right panel the sky image is shifted by and amount $\boldsymbol{\Phi} = (0, 0.3k)$. Because the shortest baseline is too long (0.8λ), the Fourier transform of the visibilities only cover up to about 0.7 in k_x and k_y, so in contrast with Figure 6-18, it is impossible to determine $\boldsymbol{\Phi}$ by merely looking at where the sky image is centered without prior knowledge of the sky.

sky.

6.B   Appendix: A Hierarchical Redundant Calibration Scheme with $\mathcal{O}(N)$ Scaling

One of the major advantages of an omniscope is its $N \log N$ cost scaling where N is the number of antennas. However, existing calibration techniques, including the ones presented in this paper, require all of the visibilities to compute the calibration parameters. Since the cost for computing the visibilities alone scales as N^2, this is a lower bound to the computational cost of existing calibration schemes regardless of the actual algorithm. While current instruments with less than 10^3 elements can afford full N^2 cross-correlation, such computation will be extremely demanding for a future omniscope with 10^4 or more elements. Thus, to take advantage of the $N \log N$ scaling of an omniscope with large N, it is necessary to have a calibration method whose
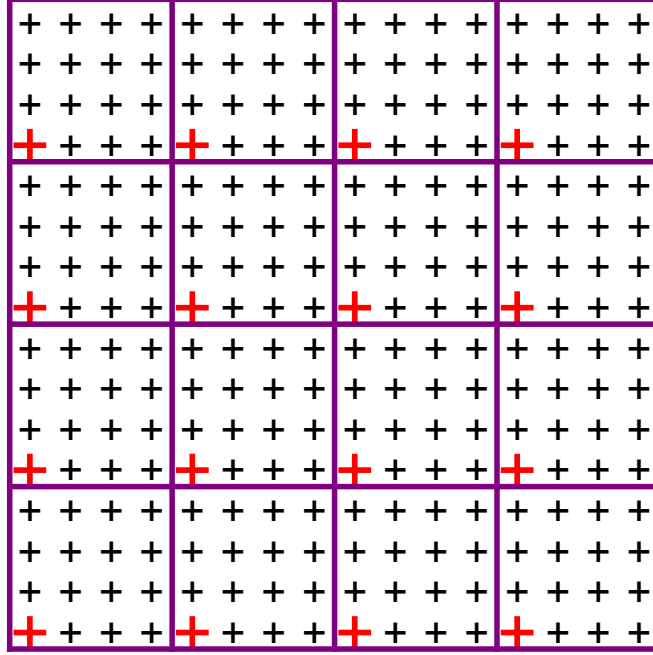
Figure 6-20: Example of the hierarchical calibration method for 256 antennas (marked by +-symbols) viewed as a 2-level hierarchy of 4×4 arrays ($m = 16, n = 2$). Our method first calibrates each sub-array independently with both relative and absolute calibrations. This produces calibration parameters for every antenna, up to one phase degeneracy ψ per sub-array. We can remove these 16 phase degeneracies among sub-arrays by choosing one antenna out of each sub-array (marked red and thick) and performing calibration on these 16 antennas. Thus we have calibrated the whole 256 antenna array by performing 16-antenna calibration 16+1=17 times. This can be generalized to a hierarchy with more levels by placing 16 such 256-antenna arrays in a 4×4 grid to get a 4096-antenna array, and then repeating to obtain arrays of exponentially increasing size. As shown in the text, the computational cost for this calibration method scales only linearly with the number of antennas.

cost scaling is less than $N \log N$. In this section, we describe a such a method using a hierarchical approach, and show that its computational cost scales only linearly with the number of antennas.

Figure 6-20 illustrates the hierarchical calibration method for an example with a 256 antennas in a 16×16 regular grid, viewed as a 2-level hierarchy of 4×4 grids. More generally, consider an n-level hierarchy with m sub-arrays at each level, containing a total of $N = m^n$ antennas; the example in Figure 6-20 corresponds to $m = 16$, $n = 2$ and $N = 256$.[30] Let B_m denote the computational cost of calibrating a basic

---

[30]It is easy to see that for a regular grid of N antennas, N need not an exact power of m to obtain

m-antenna array[31] Let C_n denote the computational cost of calibrating the entire n-level hierarchy containing all N antennas. We have $C_1 = B_m$ by definition and

$$C_{n+1} = mC_n + B_m \tag{6.26}$$

since, as explained in the caption of Figure 6-20, we can calibrate the m sub-arrays at cost C_n each and then calibrate the m reference antennas (one from each sub-array) at cost B_m. Solving this recursion relation gives

$$
\begin{aligned}
C_n &= B_m(1 + m(1 + m(1 + m(1 + \cdots)))) \\
&= B_m \sum_{k=0}^{n-1} m^k = \frac{m^n - 1}{m - 1} B_m \\
&= \frac{N - 1}{m - 1} B_m = \mathcal{O}(N).
\end{aligned}
\tag{6.27}
$$

Eq. 6.27 implies that for a fixed m, the computational cost for calibrating a 10^5 antenna array will be 10 times that of a 10^4 antenna array. Intuitively, the cost reduction comes from not computing cross-correlations among most pairs of antennas. In the simple case in Figure 6-20, only one visibility is computed between each pair of sub-arrays, rather than 256 visibilities in a full correlation scheme. Because of the reduced number of cross-correlations computed, we expect the quality of calibration parameters to be worse than that in the full correlation case. Since both the precision of calibration parameters and the computational cost depend on m, one can tune the value of m to achieve an optimal balance between precision and computational cost.

the scaling that we will derive.

[31] B_m includes the cost to compute cross-correlations between the m antennas, as well as both relative and absolute calibrations. The cost B_m is unimportant for the scaling as long as it is independent of n.

6.C Appendix: Fast Algorithm to Simulate Visibilities Using Global Sky Model

For both traditional self-calibration and the absolute calibration described in this paper, it is crucial to have accurate predictions for the visibilities. This requires simulation of both the contributions of bright point sources and diffuse emission, which can be added together due to the linearity of visibilities. While it is computationally easy to compute the contributions of point sources of known flux, it is much harder to compute visibilities from diffuse emission such as that modeled by the global sky model (GSM, de Oliveira-Costa et al. 54). Dominated by Galactic synchrotron radiation, this diffuse emission is especially important for the low frequencies and angular resolutions typical of current 21 cm experiments.

In general, we want to compute visibilities

$$y_u(f,t) = \int s(\hat{\boldsymbol{k}}, f, t) B(\hat{\boldsymbol{k}}, f) e^{i\boldsymbol{k}\cdot\boldsymbol{d}_u} d\Omega_k, \tag{6.28}$$

where $s(\hat{\boldsymbol{k}}, f, t)$ is the magnitude squared of the global sky model at time t in horizontal coordinates, and $B(\hat{\boldsymbol{k}}, f)$ the magnitude squared of the primary beam at a given frequency. Performing the integral by summing over all n_{pix} pixels in the GSM takes $\mathcal{O}(n_{\text{pix}} n_b n_f n_t)$ computational steps, where n_b is the number of unique baselines one simulates, n_f is the number of frequency bins, and n_t is the number of visibilities one simulates for one sidereal day.

The faster algorithm we describe here takes only $\mathcal{O}(n_{\text{pix}} n_b n_f)$ steps, by taking advantage of the smoothness of the primary beam as well as the diurnal periodicity in Earth's rotation. It applies only to drift-scanning instruments, so $B(\boldsymbol{k}, f, t) = B(\boldsymbol{k}, f)$ in horizontal coordinates, and is similar in spirit to the ideas proposed by Shaw et al. [197].

The key idea is to decompose Equation 6.28 as follows:

$$y_u(f,t) = \sum_{\ell,m} a_{\ell m}^f \mathcal{B}_{\ell m}^{uf} e^{im\phi(t)}, \tag{6.29}$$

344

where each $a_{\ell m}^f$ is a spherical harmonic component of the GSM at a given frequency, and each $\mathcal{B}_{\ell m}^{uf}$ is a spherical harmonic component of $B(\hat{\boldsymbol{k}}, f)e^{i\boldsymbol{k}\cdot\boldsymbol{d}_u}$, both in equatorial coordinates. In this appendix, we describe precisely how to perform this decomposition and why it decreases the computational cost of calculating visibilities from the GSM.

6.C.1 Spherical Harmonic Transform of the GSM

The GSM of de Oliveira-Costa et al. [54] is composed of three HEALPIX maps of size n_{side} describing different frequency-independent sky principal components $s^c(\hat{\boldsymbol{k}})$ and the relative weights of each component $w^c(f)$ that encode the frequency dependence. We can decompose the spatial dependence into spherical harmonics,

$$a_{\ell m}^c = \int Y_{\ell m}^*(\hat{\boldsymbol{k}}) s^c(\hat{\boldsymbol{k}}) d\Omega_k \tag{6.30}$$

in $\mathcal{O}(n_{\mathrm{pix}}^{\frac{3}{2}})$ steps, due to the advantage of HEALPIX format [79]. The frequency dependence of the spherical harmonic coefficients of the sky is given by

$$a_{\ell m}^f = \sum_c a_{\ell m}^c w^c(f), \tag{6.31}$$

and the total complexity of computing the coefficients $a_{\ell m}^f$ is $\mathcal{O}(n_{\mathrm{pix}}^{\frac{3}{2}}) + \mathcal{O}(n_f)$.

6.C.2 Spherical Harmonic Transform of the Beam and Phase Factors

Next, we would like to compute the spherical harmonics components of $B(\hat{\boldsymbol{k}}, f)e^{i\boldsymbol{k}\cdot\boldsymbol{d}_u}$:

$$\mathcal{B}_{\ell m}^{uf} = \int Y_{\ell m}^*(\hat{\boldsymbol{k}}) B(\hat{\boldsymbol{k}}, f) e^{i\boldsymbol{k}\cdot\boldsymbol{d}_u} d\Omega_k. \tag{6.32}$$

Substituting the spherical harmonic decompositions of $B(\hat{\boldsymbol{k}}, f)$ and $e^{i\boldsymbol{k}\cdot\boldsymbol{d}_u}$ gives

$$
\begin{aligned}
\mathcal{B}_{\ell m}^{uf} = & \int Y_{\ell m}^*(\hat{\boldsymbol{k}}) \sum_{\ell'm'} B_{\ell',m'}^f Y_{\ell'm'}(\hat{\boldsymbol{k}}) \\
& \times \sum_{\ell''m''} 4\pi i^{\ell''} j_{\ell''}\left(\frac{2\pi f}{c} d_u\right) Y_{\ell''m''}^*(\hat{\boldsymbol{d}}_u) Y_{\ell''m''}(\hat{\boldsymbol{k}}) d\Omega_k \\
= & \sum_{\ell'm'} \sum_{\ell''m''} 4\pi i^{\ell''} j_{\ell''}\left(\frac{2\pi f}{c} d_u\right) B_{\ell'm'}^f Y_{\ell''m''}^*(\hat{\boldsymbol{d}}_u) \\
& \times \int Y_{\ell m}^*(\hat{\boldsymbol{k}}) Y_{\ell'm'}(\hat{\boldsymbol{k}}) Y_{\ell''m''}(\hat{\boldsymbol{k}}) d\Omega_k \\
= & \sum_{\ell'm'} \sum_{\ell''m''} 4\pi i^{\ell''} j_{\ell''}\left(\frac{2\pi f}{c} d_u\right) B_{\ell'm'}^f Y_{\ell''m''}^*(\hat{\boldsymbol{d}}_u) \\
& \times \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \\
& \times (-1)^m \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ -m & m' & m'' \end{pmatrix},
\end{aligned} \tag{6.33}
$$

where $j_\ell(x)$ is the spherical Bessel function, $\ell'm'$ represent quantum numbers when expanding the primary beam, $\ell''m''$ represent quantum numbers when expanding $e^{i\boldsymbol{k}\cdot\boldsymbol{d}_u}$, and the Wigner-3j symbols are results of integrating the product of three spherical harmonics. Because the Wigner-3j symbols vanish unless $\ell - \ell' \leq \ell'' \leq \ell + \ell'$ and $-m + m' + m'' = 0$, the above sum simplifies to

$$
\begin{aligned}
\mathcal{B}_{\ell m}^{uf} = & \sum_{\ell'm'} \sum_{\ell''=\ell-\ell'}^{\ell+\ell'} 4\pi i^{\ell''} j_{\ell''}\left(\frac{2\pi f}{c} d_u\right) B_{\ell'm'}^f Y_{\ell''m''}^*(\hat{\boldsymbol{d}}_u) \\
& \times \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \\
& \times (-1)^m \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ -m & m' & m'' \end{pmatrix},
\end{aligned} \tag{6.34}
$$

where $m'' = m - m'$. Note that ℓ', m' and ℓ'' in this sum are all limited to the range of ℓ-values where the spherical harmonics components for the primary beam are non-zero, so the complexity for this triple sum is $n_{B\text{pix}}^{\frac{3}{2}}$, where $n_{B\text{pix}}$ is the number of non-zero spherical harmonics components for the primary beam. Since the cost for

each $\mathcal{B}_{\ell m}^{uf}$ is $n_{\mathrm{Bpix}}^{\frac{3}{2}}$, and there are $n_b n_f n_{\mathrm{pix}}$ of them, the computational complexity of calculating all $\mathcal{B}_{\ell m}^{uf}$-coefficients scales like $\mathcal{O}(n_b n_f n_{\mathrm{pix}} n_{\mathrm{Bpix}}^{\frac{3}{2}})$.

6.C.3   Computing Visibilities

By performing a coordinate transformation on Equation 6.28 from horizontal coordinates (corresponding to the local Horizon at the observing site) to equatorial coordinates, the time dependence of $s(\hat{\boldsymbol{k}})$ is transferred to $B(\hat{\boldsymbol{k}})$ and $\boldsymbol{d}_u$. We can now calculate $y_u(f, t)$ by computing

$$
\begin{aligned}
y_u(f, t) &= \int s(\hat{\boldsymbol{k}}) B^{ft}(\hat{\boldsymbol{k}}) e^{i\boldsymbol{k} \cdot \boldsymbol{d}_u(t)} d\Omega_k \\
&= \sum_{\ell m} a_{\ell m}^* \mathcal{B}_{\ell m}^{uft}.
\end{aligned}
\tag{6.35}
$$

Since the time dependence of $\mathcal{B}_{\ell m}^{uft}$ is a constant rotation along the azimuthal direction, we can write the above as

$$
y_u(f, t) = \sum_{\ell m} a_{\ell m}^* \mathcal{B}_{\ell m}^{uf} e^{im\phi(t)} = \sum_m c_m^{uf} e^{im\phi(t)},
\tag{6.36}
$$

where we have defined

$$
c_m^{uf} \equiv \sum_\ell a_{\ell m}^* \mathcal{B}_{\ell m}^{uf},
\tag{6.37}
$$

which can be computed in $\mathcal{O}(n_b n_f n_{\mathrm{pix}})$ steps. Given c_m^{uf}, it is clear that we can evaluate Equation 6.36 using a fast Fourier Transform (FFT), whose cost is

$$
\mathcal{O}(n_b n_f n_t \log(n_t)).
\tag{6.38}
$$

Note that this FFT in Equation 6.36 has no n_{pix} dependence, because we always need to zero-pad c_m to length n_t before the FFT. In summary, the total complexity of all

of the above steps is

$$\mathcal{O}\left(n_{\text{pix}}^{\frac{3}{2}}\right) + \mathcal{O}(n_f) + \mathcal{O}\left(n_b n_f n_{\text{pix}} n_{B\text{pix}}^{\frac{3}{2}}\right)$$

$$+ \mathcal{O}\left(n_b n_f n_{\text{pix}}\right) + \mathcal{O}(n_b n_f n_t \lg(n_t))$$

$$\approx \mathcal{O}\left(n_b n_f n_{\text{pix}} n_{B\text{pix}}^{\frac{3}{2}}\right). \tag{6.39}$$

This does not scale with n_t, unlike the naive integration's $\mathcal{O}(n_b n_f n_{\text{pix}} n_t)$. Thus with a spatially smooth beam whose $n_{B\text{pix}} \ll n_t^{\frac{2}{3}}$, the algorithm described here is much faster than the naive numerical integration approach described at the beginning of this Appendix.

Part III

The Cosmic Dawn on the Horizon

Chapter 7

Detecting the 21 cm Forest in the 21 cm Power Spectrum

The content of this chapter was submitted to the Monthly Notices of the Royal Astronomical Society *on October 30, 2013 and published [63] as* Detecting the 21 cm forest in the 21 cm power spectrum *on May 20, 2014.*

7.1 Introduction

Observations of emission and absorption at 21 cm from the neutral intergalactic medium (IGM) at high redshift will offer an unprecedented glimpse into the cosmic dark-ages up through the epoch of reionization (EoR), constraining both fundamental cosmological parameters and the properties of the first stars and galaxies [74, 154, 188, for reviews]. Direct mapping of the 21 cm signal during the EoR is likely a decade or more away, requiring projected instruments such as the Square Kilometer Array (SKA). However, a first generation of experiments attempting to detect the power spectrum are already underway. These include the Low Frequency ARray [244, LOFAR], the Murchison Widefield Array [220, MWA], the Precision Array for Probing the Epoch of Reionization [173, PAPER], and the Giant Metre-wave Telescope [167, GMRT]. The MWA, PAPER, and LOFAR have the potential to achieve statistical detections of brightness temperature fluctuations within the next several years [25,

139, 15, 147].

Most theoretical investigations of observing neutral hydrogen in the EoR have focused on IGM emission and absorption against the Cosmic Microwave Background (CMB). It has also been recognized by Carilli et al. [34], Furlanetto and Loeb [70], Xu et al. [243], Mack and Wyithe [132], Ciardi et al. [43] that the 21 cm forest, HI absorption in the spectra of background radio-loud (RL) active galactic nuclei (AGN), can be used to probe the IGM's thermal state.

Studies of the forest have focused on its detection in the frequency spectra of a known RL source to glean information on the thermal properties of the absorbing IGM. The possibility for such a study depends on the existence of high redshift RL sources. As of 2013, the RL source distribution is only well constrained out to $z \sim 4$ (see de Zotti et al. [55] for review). Theoretical work suggests that at 100 MHz hundreds of $S \sim 1$ mJy sources with redshifts greater than 10 might exist within one of the $(30°)^2$ fields of view (FoV) offered by existing and upcoming wide field interferometers [85, 236] (hereafter H04 and W08 respectively). However the discovery of a suitable source at high redshift entails an extensive follow up program to measure photometric redshifts of radio selected candidates.

Should sufficiently RL sources exist, a line of sight (LoS) detection of individual absorption features will require large amounts of integration time on a radio telescope with the collecting area comparable to the Square Kilometer Array (SKA). At reionization redshifts, [132] find that a 5σ detection of an individual absorption feature with a $z \approx 9$ Cygnus A type source[1] would require years of integration on an SKA-like instrument. Ciardi et al. [43] find that after 1000 hours of integration only 0.1% of the LoS in an IGM simulation box contained regions of large enough optical depth to produce absorption features[2] observable by LOFAR. Hence a detection of the forest with a present day interferometer would require a very rare juxtaposition of an extremely loud RL source with an outlying optical depth feature. Even if this detection were achieved, it is unlikely that significant inferences on the thermal history could

---

[1] flux density at 151 MHz of $S_{151} \approx 20$ mJy and spectral index of $\alpha \approx 1.05$

[2] against a $S_{129} \approx 50$ mJy source at $z \sim 7$

be made from only a handful of such observations.

While detecting individual absorption features presents an enormous challenge, statistical methods have been demonstrated to reduce the necessary integration times. One target for a statistical detection is the increased variance in flux, along the LoS. It is shown in [132] that the integration time required for detecting this variance increase for a Cygnus A source, is only a few weeks with an SKA-like telescope, as apposed to the decades needed for detecting a single feature. Ciardi et al. [43] find that LOFAR could detect the global suppression in the spectrum of a 50 mJy source at $z \sim 12$ with a 1000 hour integration, though they note that a detection by LOFAR is unlikely due to excessive RFI in the FM band (80 MHz $\lesssim \nu \lesssim$ 108 MHz).

The possibility of a statistical detection of the forest using information from the wide FoV available to the current and upcoming generations of experiments has not yet been investigated. Observing the forest signature in the 21 cm power spectrum would integrate the signal from many high redshift sources within a FoV, reducing the sensitivity requirements of the instrument. Also, a power spectrum detection does not require a priori knowledge of high redshift sources. Hence the technique we describe can put constraints on both the properties of the IGM, such as the heating and reionization history, and the population of high redshift RL sources. It is likely that 21cm forest absorption features could be fruitfully explored using high-order statistical measures as well, but we do not consider those in this paper.

In this proof-of-concept, we begin to explore the characteristics and observability of the forest in the 21 cm power spectrum. We derive analytically the features that the global forest should introduce to the power spectrum and confirm their existence by combining semi-numerical simulations of the IGM, computed with 21cm-FAST [144], with the semi-empirical model of the high redshift population of RL sources from W08. We find that in all heating scenarios studied, the contribution to the 21 cm fluctuations by the absorption of our RL sources is comparable to or dominates the contribution from the brightness temperature on small spatial scales ($k \gtrsim 0.50$ Mpc^{-1}). To determine the detectability of the forest in the power spectrum, we perform sensitivity calculations for several radio arrays with designs similar to the

MWA, including a future array with a collecting area of $\sim 0.1 \text{km}^2$, similar to the planned Hydrogen Epoch of Reionization Array (HERA). In order to give the reader a sense of how the strength of this signal scales across a large range of of radio loud source populations, we extrapolate the expected S/N of the Forest using our analytic expression for the signal strength.

This paper is organized as follows. In Section 7.2 we provide the theoretical background and use a toy model to derive the morphology of the 21 cm forest power spectrum; relating its shape and amplitude to the optical depth power spectrum and the radio luminosity function. In Section 7.3 we describe the semi-numerical simulations of the IGM along with the semi-empirical RL source distribution of W08 and how we combine them to simulate the wide field forest. In Section 8.3 we discuss our results and identify the separate regions of k-space that may be used to independently constrain the thermal history of the IGM and the high redshift RL distribution. In Section 7.5 we explore the prospects for detecting the forest in spherically averaged power spectrum measurements considering the sensitivity of current and future radio arrays. In Section 7.6 we extrapolate our detectability results across a broad range of source populations and X-ray heating scenarios.

Throughout this work we assume a flat universe with the cosmological parameters $h = 0.7$, $\Omega_\Lambda = 0.73$, $\Omega_M = 0.27$, $\Omega_b = 0.082$, $\sigma_8 = 0.82$, and $n = 0.96$ as determined by the WMAP 7-year release [111]. All cosmological distances are in comoving units unless stated otherwise.

7.2 Theoretical Background

In this section we establish our notation and present a basic mathematical description of how forest absorption modifies the 21 cm brightness temperature signal.

7.2.1 Notation

We adopt the Fourier transform convention

$$\widetilde{f}(\mathbf{k}) = \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}). \tag{7.1}$$

In addition, we often refer to cylindrical Fourier coordinates where $k_\perp \equiv \sqrt{k_x^2 + k_y^2}$ and $k_\parallel \equiv |k_z|$. The power spectrum of a field A over a comoving volume V is defined as

$$P_A = \frac{1}{V}\langle|\widetilde{\Delta A}|^2\rangle \tag{7.2}$$

and the cross power spectrum between fields A and B over V is given by

$$P_{A,B} = \frac{1}{V}\langle\widetilde{\Delta A}\widetilde{\Delta B}^*\rangle \tag{7.3}$$

where

$$\Delta A = A - \langle A\rangle \tag{7.4}$$

and $\langle A\rangle$ is defined as the ensemble average of A though in practice it is computed by averaging over some spatial or Fourier volume. In our discussion, we will also be referring to the one dimensional LoS power spectrum (not to be confused with the 1D spherical power spectrum) of a field A along a LoS column of comoving length L.

$$P_A^{LoS}(k_z) = \frac{1}{L}\int dz dz' \Delta A(z)\Delta A(z')e^{ik_z(z-z')} \tag{7.5}$$

Finally, we use Δ^2 to denote the dimensionless power spectrum

$$\Delta^2(k) \equiv \frac{k^3}{2\pi^2}P(k) \tag{7.6}$$

7.2.2 The Forest's Modification of the Brightness Temperature

The forest absorption traces the optical depth of the IGM and will therefore introduce a signal on similar spatial scales as the 21 cm brightness temperature. We now discuss this signal in detail. The optical depth of a high redshift HI cloud is given by [74]

$$\tau_{21} \approx .0092(1+\delta)(1+z)^{3/2}\frac{x_{HI}}{T_S}\left[\frac{H(z)/(1+z)}{dv_{\parallel}/dr_{\parallel}}\right]. \tag{7.7}$$

δ is the fractional baryonic over-density, H(z) is the Hubble factor, $dv_{\parallel}/dr_{\parallel}$ is the velocity gradient along the LoS (including the Hubble expansion), and x_{HI} is the neutral hydrogen fraction. The numerical factor in front of Equation (7.7) is computed from fundamental constants and is independent of cosmology. The spin temperature, T_s is defined by the relative population densities of the two hyperfine energy levels, n_1 and n_0 [66]

$$\frac{n_1}{n_0} = 3\exp\left(-\frac{h\nu_{21}}{k_B T_s}\right). \tag{7.8}$$

Where, h is Plank's constant, k_B is the Boltzmann constant, and $\nu_{21} = 1420.41\text{MHz}$ is the rest frame frequency of the hyperfine transition radiation.

Prior works on 21 cm tomography assume that the sky temperature at $\nu = \nu_{21}/(1+z)$ in the direction of an HI cloud is given by

$$T_{sky} = \frac{T_s}{(1+z)}(1-e^{-\tau_{21}}) + \frac{T_{CMB}}{(1+z)}e^{-\tau_{21}} + T_{fg}. \tag{7.9}$$

where T_{CMB} is the comoving temperature of the cosmic microwave background radiation and T_{fg} is the temperature of foreground emission including synchrotron radiation of the Galaxy, resolved point sources, free-free emission, and radio emission from unresolved point sources below the confusion limit [57, 107, 54, 231].

The first term in Equation (7.9) includes both the 21 cm emission and self absorption of the HI cloud, hence it is multiplied by a factor of $(1-e^{-\tau_{21}})$. The second term describes the observed intensity of a background source shining through the cloud so

356

its temperature is attenuated by $e^{-\tau_{21}}$. The third term describes radiation emitted by sources closer than the cloud so its intensity unaffected by τ_{21}.

21 cm experiments seek to measure the difference between the first two terms of Equation (7.9) and T_{CMB}. This difference is often referred to as the "differential brightness temperature" and is given by [74]

$$T_b = \frac{(T_s - T_{CMB})}{(1+z)}(1 - e^{-\tau_{21}}) \approx \frac{T_s - T_{CMB}}{(1+z)}\tau_{21}. \tag{7.10}$$

We depart from previous work by considering the effect of radio loud sources behind the HI cloud whose combined observed[3] brightness temperature we denote as T_{RL}. Including these background sources, Equation (7.9) becomes

$$T'_{sky} = \frac{T_s}{(1+z)}(1 - e^{-\tau_{21}}) + \frac{T_{CMB}}{(1+z)}e^{-\tau_{21}} + T_{RL}e^{-\tau_{21}} + T_{fg} \tag{7.11}$$

T_{fg} and T_{RL} are expected to have predominantly smooth spectra which reside within a limited region of Fourier space known as the "wedge" [50, 156, 230]. Smooth spectrum components may be removed by filtering[172] or subtraction [28, 122, 59], both employing the separation of the foregrounds and signal in the Fourier domain.

We will focus on the fluctuating signal, assuming that the smooth spectrum components of the foregrounds and background sources are properly avoided and/or subtracted. The effective differential brightness temperature now includes a contribution from the forest absorption features.

$$T_b \to T'_b \approx T_b - T_f \tag{7.12}$$

where $T_f = T_{RL}\tau_{21}$ is the "forest temperature". We can see how the power spectrum is transformed by the inclusion of T_f by inserting Equation (7.12) into Equation (7.2)

$$P_b \to P'_b = P_b + P_f - 2\mathrm{Re}(P_{f,b}) \tag{7.13}$$

---

[3]In accordance with much of the literature, we use the observed temperature for T_{RL} and T_{fg}, rather than the comoving temperature as we have for T_s and T_{CMB}. As a result, there are no factors of $(1+z)$ under T_{RL} or T_{fg}.

Where $P_b \equiv P_{T_b}$, $P_b' \equiv P_{T_b'}$, $P_f \equiv P_{T_f}$ and $P_{f,b} \equiv P_{f,T_b}$. Equation (7.13) sums up how the forest modifies the power spectrum that we expect to observe in upcoming 21 cm observations. Essentially, smooth spectrum power from T_{RL} is leaked from the largest spatial modes to those occupied by T_b via a convolution with the power spectrum of the optical depth field. The magnitude of this leakage will increase with the magnitude of the optical depth.

7.2.3   The Morphology of the Forest Power Spectrum

The first thing one might ask concerning the forest contribution described in Equation (7.13) is how the magnitudes of the two contributions compare to each other and what their qualitative features are. While we will answer these questions with simulations it is useful to gain as much insight as we can through analytic methods. We start with P_f which can be decomposed (see Appendix 7.A for a derivation) into a sum of auto power spectra P_j originating from each individual RL source behind or within an imaged volume of IGM and their cross power spectra, $P_{j,k}$.

$$P_f = \frac{1}{V}\left\langle \left|\widetilde{T_{RL}\tau_{21}}\right|^2\right\rangle = \sum_j P_j + 2\mathrm{Re}\left(\sum_{j<k} P_{j,k}\right) \tag{7.14}$$

If all of the background sources are unresolved[4] then each P_j is the absolute magnitude of the Fourier transform of a function that is a delta function in the perpendicular to LoS directions. As a result, each P_j in Equation (7.14) is constant in $k_\perp$. The cross multiplying $P_{j,k}$ terms are not so simple; however, we show in Appendix 7.A that in the absence of clustering, the cross sum only contributes to P_f at the 10% level for $k_\parallel \gtrsim 0.1$ Mpc^{-1}. At these scales, P_f only has considerable structure along $k_\parallel$

$$P_f(\mathbf{k}) \approx \sum_j P_j(k_\parallel) = \frac{D_M^2 \lambda^4}{4 k_B^2 \Omega_{cube}} P_{\tau_{21}}^{LoS}(k_\parallel)\langle \sum_j s_j^2 \rangle \tag{7.15}$$

---

[4]a fair assumption given the large synthesized beams of interferometers and small angular extent of high redshift sources

where $\lambda = \lambda_{21}(1+z)$ is the observed wavelength of 21 cm light emitted from the center of the imaged volume, D_M is the comoving distance to the data cube, and Ω_{cube} is the solid angle subtended by the cube. In the second step, we have expressed each P_j in terms of the flux of each source, s_j, and the 1D power spectrum along the line-of-site to that source, $P_{\tau_{21}}^{LoS}$. In addition P_f is positive so that it will always add to the power spectrum amplitude

We can convert the sum in Equation (7.15) to an integral over the radio luminosity function

$$P_f \approx \frac{cD_M^2\lambda^4}{4k_B^2}P_{\tau_{21}}^{LoS}(k_\parallel)\int_z^\infty\int_0^\infty s'^2\rho(z,z',s')\frac{D_M^2(z')}{H(z)}dz'ds' \qquad (7.16)$$

where $\rho(z,z',s') \equiv \frac{dN}{ds'dV_c}$ is the differential number of radio loud sources per comoving volume at redshift z' per flux bin at observed frequency $\nu_{21}/(1+z)$ and s' is the flux at $\nu = \nu_{21}/(1+z)$.

Equation (7.16) tells us that the amplitude of the forest power spectrum is set by the integral over the high redshift radio luminosity function multiplied by the average optical depth squared[5] while the shape of the forest power spectrum is set by the 1D LoS power spectrum of optical depth fluctuations.

$P_{f,b}$ does not separate so conveniently but we can gain insight into whether it adds or subtracts to Equation (7.13) by considering the physical phenomena that govern T_f and T_b. Expanding Equation (7.10) and T_f in terms of the IGM properties using Equation (7.7) one can see that $P_{f,b}$ is the cross power spectrum between the two quantities:

$$T_b \approx 9x_{HI}(1+\delta)(1+z)^{1/2}\left[1-\frac{T_{CMB}}{T_s}\right]\left[\frac{H(z)}{dv_\parallel/dr_\parallel}\right] \text{ mK} \qquad (7.17)$$

and

$$T_f \approx 0.009x_{HI}(1+\delta)(1+z)^{1/2}\frac{T_{RL}}{T_s}\left[\frac{H(z)}{dv_\parallel/dr_\parallel}\right] \qquad (7.18)$$

Before the reionization era, x_{HI} is relatively homogenous so that fluctuations in T_b are governed primarily by those in T_s. Regions of the IGM with larger T_s will

---

[5]By our definition, the power spectrum is the Fourier transform squared of $\Delta\tau_{21}$, not $\delta_{\tau_{21}} = \Delta\tau_{21}/\langle\tau_{21}\rangle$ which is often used in other work. Hence our power spectrum amplitude is set by $\langle\tau_{21}\rangle^2$

have more positive T_b but smaller T_f. Because of this anti-correlation between T_f and T_b, $\mathrm{Re}(P_{f,b})$ is negative during the pre-reionization era and the net effect will be for it to increase the power spectrum amplitude through its negative contribution in Equation (7.13). At lower redshifts, after X-rays have heated the IGM, $T_s \gg T_{CMB}$, and T_b becomes independent of T_s. As a result, T_b is always positively correlated with x_{HI} as is T_f. $\mathrm{Re}(P_{f,b})$ is positive with a net effect of subtracting from the power spectrum amplitude. We are unable to make any more progress analytically, but we will reexamine the cross power term in our simulation results below.

We now move on to describe our simulations. We will return to our discussion of the power spectrum morphology in the context of our simulation results in Section 8.3.

7.3 Simulations

In this section we describe the semi-numerical simulations that we use to explore a range of IGM thermal histories along with the the semi-empirical RL source model that we employ to add the 21 cm forest signal.

7.3.1 Simulations of the Optical Depth of the IGM

Our IGM simulations are run using a parallelized version of the public, semi-numerical 21cmFAST code[6] described in Mesinger et al. [144]. Tests of the code can be found in Mesinger and Furlanetto [143], Zahn et al. [245], Mesinger et al. [144]. The simulation box is 750 Mpc on a side, with resolution of 500^3. Different scenarios for τ_{21} can be obtained by exploring histories of the spin temperature, T_s and/or the neutral fraction, x_{HI}.

21cmFAST includes sources of both UV ionizing photons and X-rays. The former dominate reionization (i.e. x_{HI}), except for extreme scenarios we do not consider in this work [69, 138, 146]. Since a full parameter study is beyond the scope of this work, and since the bulk of the relevant signal is likely during the pre-reionization

---

[6]`http:/homepage.sns.it/mesinger/Sim`

epoch, we fix the ionizing emissivity of galaxies (and hence the reionization history), to agree with the Thompson scattering optical depth from WMAP [111]. Instead we focus on the X-ray emissivity and its impact on T_s.

T_s is affected by a variety of processes. These include Ly-α photons which couple to the hyperfine transition through the Wouthuysen-Field effect [238, 66], particle collisions, and emission or absorption of CMB photons. The coupling of T_s to these processes is described by (e.g. Furlanetto et al. 74):

$$T_s^{-1} = \frac{T_{CMB}^{-1} + x_c T_k^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha} \tag{7.19}$$

where T_k is the kinetic temperature of the HI gas, T_c is the color temperature of Ly-α photons, and x_c and x_α are the collisional and Ly-α coupling constants. Due to the high optical depth of the neutral IGM to Ly-α photons, the color temperature is very closely coupled to the kinetic temperature, $T_c \approx T_k$ [238, 93] .

Although the self-annihilation of some dark matter candidates can contribute significantly [227], in fiducial models T_k is predominantly determined by X-ray heating (e.g. Furlanetto et al. 74). Hence, we explore a range of optical depth histories by running simulations for different galactic X-ray emissivities.

We use the fiducial model of X-ray heating described in Mesinger et al. [146], adopting a spectral energy index of $\alpha = 1.5$ and an obscuration threshold of 300 eV. We parameterize the X-ray luminosity by a dimensionless efficiency parameter, f_X. Our fiducial model, $f_X = 1$ corresponds to 0.2 photons per stellar baryon, or a total X-ray luminosity above $h\nu_0 = 0.3$ keV of $L_{X,0.3+keV} \approx 10^{40}$ erg s^{-1} $(M_\odot$ yr$^{-1})^{-1}$. This choice is consistent with (a factor of $\sim$2 higher than) an extrapolation from the 0.5–8 keV measurement of Mineo et al. [148] that $L_{X,0.5-8keV} \approx 3 \times 10^{39}$ erg s^{-1} $(M_\odot$ yr$^{-1})^{-1}$.

Summarized in Table 7.1 are our three values of f_X: a "fiducial IGM" model with $f_X = 1$ corresponding to the fiducial value in Mesinger et al. [146], a "hot IGM" model with $f_X = 5$, and a "cool IGM" model with $f_X = 0.2$. In Figure 7-1 We show the evolution of the mean spin and brightness temperatures from our simulations. Over

Number Name	f_X
Hot IGM	5.0
Fiducial IGM	1.0
Cool IGM	0.2

Table 7.1: IGM Heating Parameters.
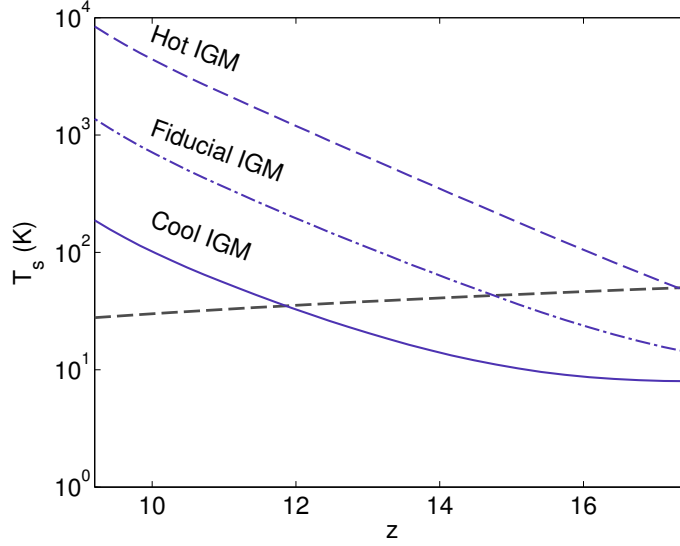
Figure 7-1: The mean thermal evolution of our IGM simulations for our three models. "cool IGM"- solid lines, "fiducial IGM"- dashed-dotted lines, and "hot IGM" - dashed lines. $\langle T_s \rangle$ is plotted in lavendar. Varying f_X effectively shifts $\langle T_s \rangle$ in redshift.

the range of emissivities considered, the effect of varying f_X is to shift the evolution of $\langle T_s \rangle$ in redshift. Because P_f varies as $\langle \tau_{21} \rangle^2 \sim \langle T_s \rangle^{-2}$ and f_X simply shifts $\langle T_s \rangle$ in redshift, this relatively modest spread in f_X is sufficient to understand a broader range of expected outcomes, as we shall see below.

7.3.2 The Model of the Radio Loud Source Distribution

We now review present constraints on the RL source distribution and describe the semi-empirical radio luminosity function that we use to simulate the global 21 cm forest. To gain perspective of how our choice of population model might compare to other theoretical work we determine which flux ranges are relevant to the sum in Equation (7.15) and compare the counts of sources in W08 to those in H04. We also describe our method for combining the simulated radio sources with our simulations

of the IGM.

7.3.2.1 Review of Constraints and Predictions of High Redshift Radio Counts

Constraints on the luminosity function of the most luminous radio loud sources are presently limited to $z \sim 4$ [55] Confirmed in these works, is that the comoving density of ultra steep spectrum sources peaks at $z \sim 2$ with little evidence for evolution out to $z \gtrsim 4.5$.

To model the abundance of RL quasars with $6 \lesssim z \lesssim 20$ one must rely on theoretical extrapolations. Haiman et al. [85] give estimates of source counts by assigning black hole masses to a halo mass function using the black hole mass-velocity dispersion relation of Wyithe and Loeb [240]. The RL fraction is derived assuming Eddington accretion, and the RL-i band luminosity correlation observed by Ivezić et al. [101].

More sophisticated attempts at predicting the bolometric luminosities of high redshift quasars up to $z = 11$ have been undertaken using hydrodynamic simulations with self consistent models for black hole growth and feedback [56]. Even with a more nuanced treatment of the luminosity distribution, the RL fraction at high redshift still remains a wide open question. Indeed, the purpose of this work is to propose a technique for determining this population by showing that an empirically motivated RL population can have significant and observable features in the power spectrum for a range of thermal scenarios.

7.3.2.2 Our Choice of Population Model

We choose to work with the RL AGN population described in W08 in which sources are generated by sampling extrapolated radio luminosity functions biased to structure from a CDM simulation. Specifically, the radio luminosity function used is that "Model C." from Willott et al. [235] which describes the high and low luminosity populations of AGN as Schechter functions. The redshift evolution of the low luminosity population is modeled as a power law in redshift while the high luminosity
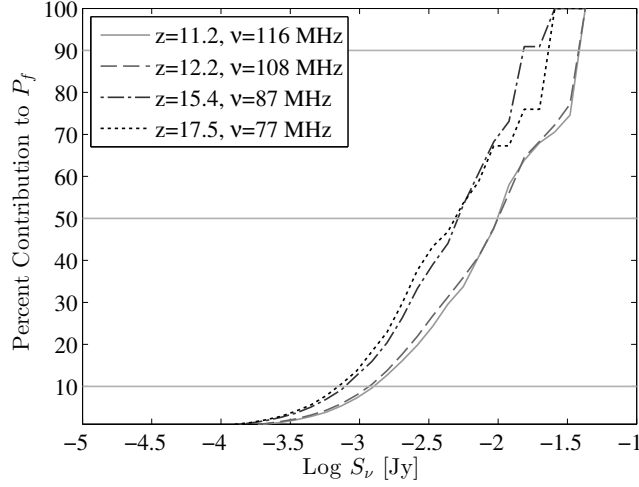
Figure 7-2: The 21 cm forest is dominated by sources in the 1-10 mJy flux range. We plot the sum of fluxes squared, in Equation (7.16), for $S < S_\nu$. A detection of P_f would constrain the high redshift source counts at these flux intervals.

component as a gaussian with a mean of $z \approx 1.9$. Lists of source positions, fluxes, and morphologies from the Wilman simulation are downloadable through a web interface[7].

Having chosen our population model, we can employ our formalism from Section 7.2 to understand which sub population of the luminosity function contributes most to P_f. In Figure 7-2, we plot the percent contribution of sources below a threshold, S_ν, to P_f from the flux squared sum in Equation (7.15). One can see that roughly 75% of the contribution to P_f comes from sources with fluxes between $1 - 10$ mJy at $80 - 115$ MHz. At lower redshifts, the integral curves are increasingly dominated by higher fluxes as the sources with the greatest fluxes increase in number. The detection or lack of detection of the features we find using this simulation would either confirm or reject the W08 model for sources with S_ν between 1 and 10 mJy. While this paper is a study of observability for one model, in future work we will determine what range of RL population this technique can constrain.

It is worth getting an order of magnitude idea of how our choice of the W08 semi-empirical model might compare to other theoretical predictions of the radio luminosity function. In Appendix 7.B we compare the source counts in our semi-empirical prediction to the more physically motivated bottom up model in H04. The

---

[7]http://s-cubed.physics.ox.ac.uk/s3_sex

counts of W08 sources contributing to the bulk of P_f tend to be more numerous than those in H04 by a factor of ≈ 10 at $z \sim 12$ to ≈ 80 for $z \sim 15 - 20$, underscoring the need for a full parameter space study. Even though such a study is beyond the scope of this paper, our extrapolated results in Figure 7-14 show that the range of populations that the power spectrum can constrain depends heavily on the IGM's thermal history.

7.3.3 Adding Sources to the Simulation

We simulate the theoretical power spectra accessible to upcoming observations by drawing 36 random sub-fields from the W08 simulations and combining them with 36 random 8MHz slices from our IGM simulations. The number of subfields is chosen to roughly correspond to the $\sim (30°)^2$ FoV of the MWA.

While our analytic approach in Section 7.2 does not account for sources within the imaged volume, we incorporate them into our simulation by determining the location of DM halos down to masses of $5 \times 10^9 M_\odot$ through the excursion-set + perturbation theory approach outlined in Mesinger and Furlanetto [143]. We then populate these dark matter halos with RL sources, monotonically assigning the most luminous sources at 151 MHz[8] to the most massive halos. Sources falling behind the cubes retain their original positions. All W08 sources are unresolved in our IGM simulation; hence, for each pixel the fluxes for all sources behind that pixel are summed together to give S_{pix}. This flux cube is converted to temperature using the Rayleigh-Jeans equation,

$$T_{pix} = \frac{\lambda^2 S_{pix}}{2 k_B \Omega_{pix}}, \tag{7.20}$$

where Ω_{pix} is the solid angle subtended be each simulation pixel[9]. Finally we introduce quasar absorption by multiplying this source cube by our τ_{21} cube $T_f \approx T_{pix} \tau_{21}$.

---

[8]We order sources by their luminosity at *observed* frequency of 151 MHz at regardless of their redshift which varies very little over the span of an 8 MHz data cube so that we are approximately comparing their rest frame luminosities.

[9]We show in Appendix 7.A, that the choice of pixel solid angle does not effect P_f

7.4 Simulation Results

In this section we present the results of our combined IGM-RL population model by computing the spherical and cylindrical power spectrum, $P(k)$, averaged over our 36 sub-cubes. We identify the regions of k-space in which the forest is dominant and might be used to constrain the high redshift radio luminosity function and discuss the morphology of the observed power spectra, verifying the essential results of Section 7.2.

7.4.1 Computing Power Spectra

Power spectra are computed using a direct Fast Fourier transform of each data cube multiplied by a kaiser window along the LoS with attenuation parameter $\beta = 3.5$. In averaging over bins of our spherical power spectra, we exclude the "wedge", the region of k-space heavily contaminated by foregrounds given by [230, 156]

$$k_\parallel \leq \sin\frac{\Theta}{2}\left(\frac{D_M(z)}{D_H}\frac{E(z)}{(1+z)}\right)k_\perp \tag{7.21}$$

where z is the redshift of a data cube's center frequency, $D_M(z)$ is the comoving distance, $E(z) = H(z)/H_0$, and Θ is the FWHM of the primary beam which we calculate using a short dipole model of the MWA antenna element. Table 7.2 gives the FWHM value of our primary beam model for several different frequencies.

7.4.2 Simulation Output and the Location of the Forest in k-space

We now discuss the power spectra output by our simulations and the significant features produced by the forest.

To isolate the the effect of the forest and to compare its significance to the brightness temperature power spectrum, P_b, we plot the fractional difference between P_b', the power spectrum with the forest ,and P_b in Figure 7-3. We see that the forest introduces a significant feature, especially at the smallest scales. This feature is most
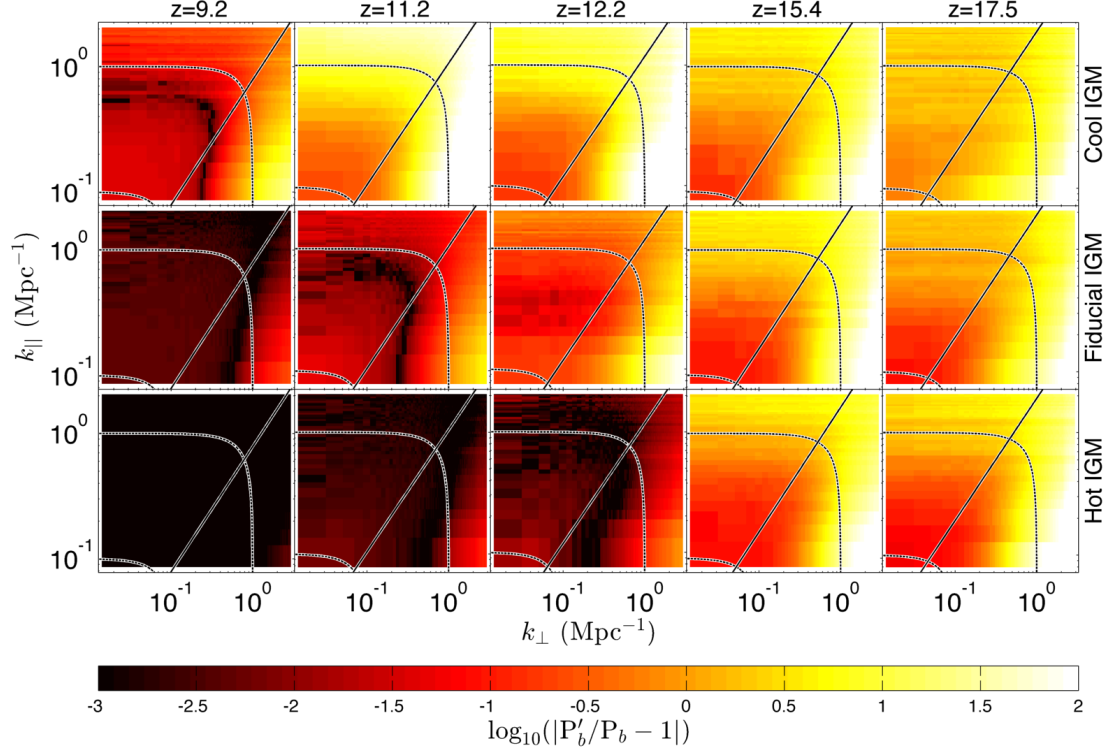
Figure 7-3: For every heating scenario we study, there is some redshift and region within the EoR window for which 21 cm forest dominates the power spectrum. Here we show the fractional difference between the power spectrum with, (P'_b), and without (P_b) the forest for the redshifts (top to bottom) 9.2, 11.2, 12.2, 15.4, and 17.5. The diagonal lines denote the location of the "wedge". By $z \gtrsim 12.2$ there is a substantial region ($k_\parallel \gtrsim 0.5$ Mpc^{-1}) of the Fourier volume that our simulations cover in which the forest dominates P_b by a factor of a few.

prominent at high redshifts and less emissive heating models, when the IGM is cool. For our cool model, the forest feature dominates P_b by over a factor of 100 for a wide range of redshifts. In the fiducial model, the dominant region is primarily at larger values of $k_\parallel$, though dominance by a factor of a few is visible at $z = 12.2$ and $z = 17.5$. In our hot model, a significant feature is visible only for $z \gtrsim 12.2$.

For all heating scenarios, there are redshifts $z \gtrsim 12.2$ in which the same region of Fourier space contains a strong forest signal that dominates P_b by a factor of at least a few. Fortunately for those interested in the brightness temperature signal, the region $k \lesssim 10^{-1}$ Mpc^{-1} remains dominated by P_b. Hence at pre-reionization redshifts, $k \lesssim 10^{-1}$ Mpc^{-1} can still be used to constrain cosmology and the thermal history of the IGM. With the thermal properties of the IGM determined, one may constrain the high redshift RL population using the forest power spectrum signal at $k \gtrsim 0.5$ Mpc^{-1}.

The first generation of interferometers will not be sensitive enough to measure the cylindrical power spectrum with high S/N but will rather measure the spherically averaged power spectrum. We compute spherically averaged power spectra from data cubes with and without the presence of forest absorption and excluding the wedge. We plot these power spectra in Figure 7-4. In all of the heating scenarios considered, the forest introduces significant power at $k \gtrsim 0.5$ Mpc^{-1} for $z \gtrsim 15.4$. Hence, it is in principle possible to constrain the distribution of RL AGN at high redshift for a range of heating scenarios.

We note that the high-k region extends into our simulations' Nyquist frequency of 2.1 Mpc^{-1}. We ensure that the forest dominance is not an aliasing effect by running simulations on a 125 Mpc cube with six times higher resolution. The results in the the overlapping k-space regions agree well with these larger volume, lower resolution simulations.

7.4.3 The Morphology of the Simulation results.

We now explain the morphology of our simulation results and verify our analytic predictions in Section 7.2.

We noted in Figure 7-3 that the 21 cm forest dominates the power spectrum both
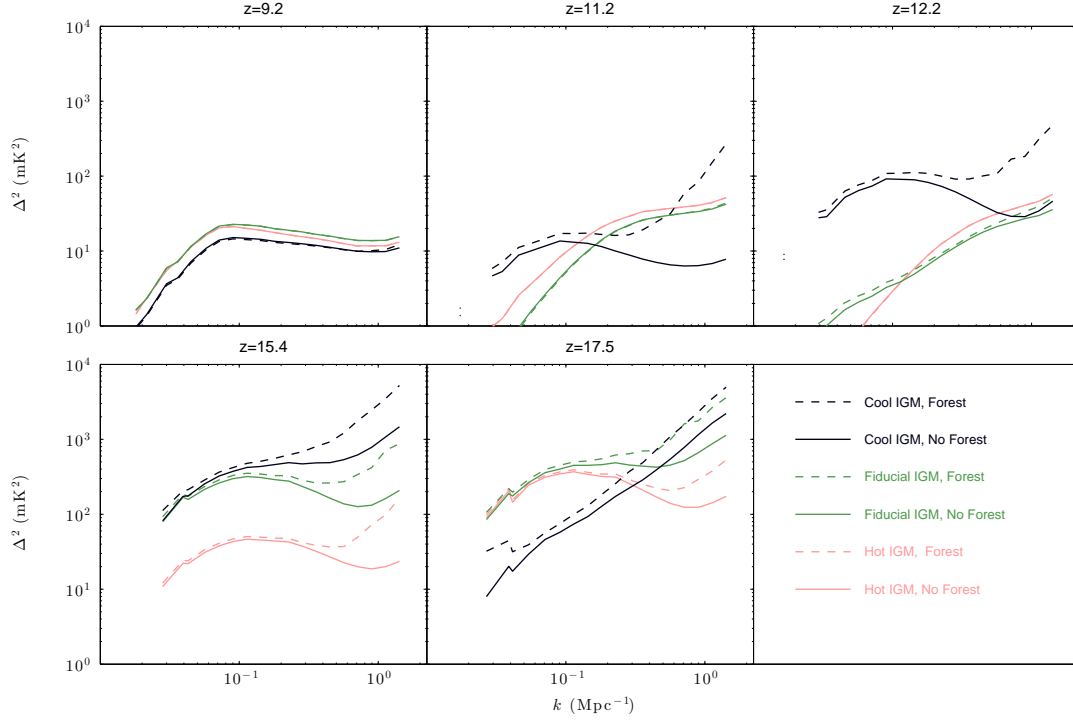
Figure 7-4: The 21 cm forest dominates the spherically averaged power spectrum for $k \gtrsim 0.5 \text{ Mpc}^{-1}$. Plotted is the spherically averaged power spectrum with (dashed lines) and without (solid lines) the presence of the 21 cm forest. In our cool model, the forest causes a significant power increase at $k \gtrsim 0.5 \text{ Mpc}^{-1}$ at redshifts as low as $z = 11.2$. At $z = 15.4$ we see a significant feature in all thermal scenarios. Our cool IGM model experiences a reduction in the power spectrum amplitude at $z \gtrsim 17.5$ as it passes through the X-ray heating peak.
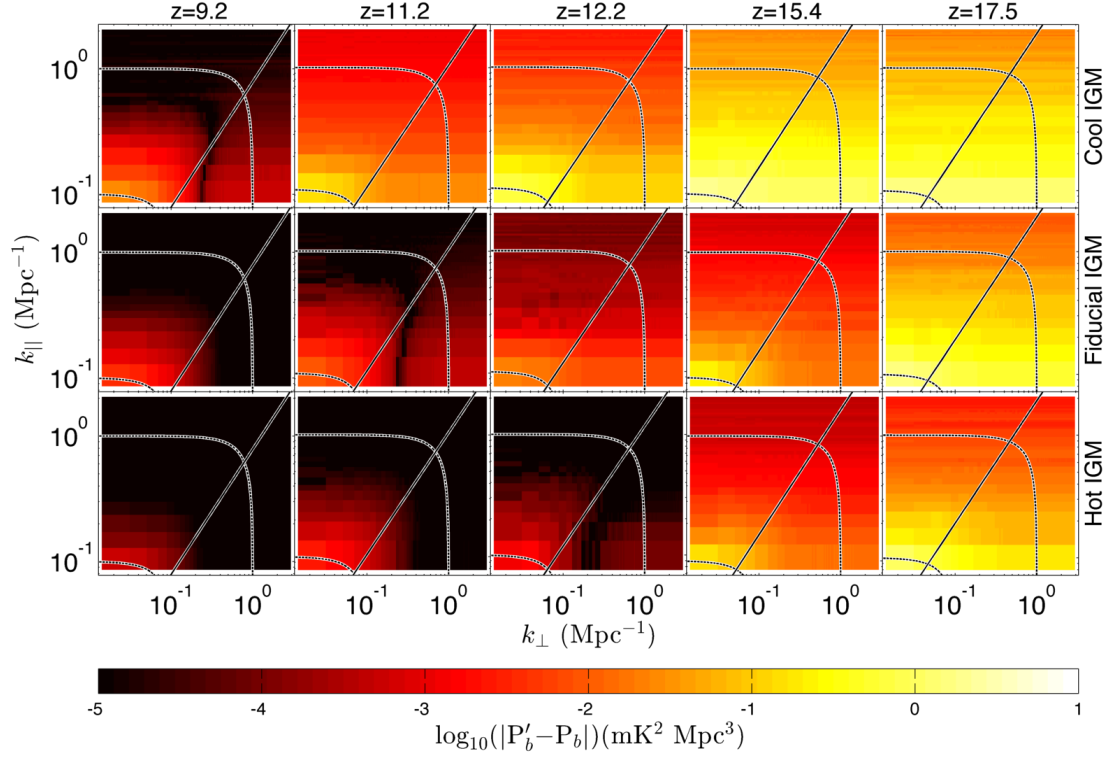
Figure 7-5: We plot the magnitude of the difference between the 21 cm power spectrum with and without the presence of the 21 cm forest including the auto-power and cross power terms of Equation (7.13). At high redshifts and low f_X, there is little $k_\perp$ structure in $P_b' - P_b$, indicating that P_f is the significant contributer. At lower redshifts and higher f_X, we see signficant $k_\perp$ structure, indicating that in a heated IGM, $P_b' - P_b$ is dominated by $P_{f,b}$ which is somewhat spherically symmetric and negative at large k. The trough in the low redshift plots marks the region where $P_f - 2\mathrm{Re}(P_{f,b})$ transitions from negative (for small k) to positive (for large k).

at large $k_\perp$ and $k_\parallel$. The former observation is consistent with a forest power spectrum that is uniform in $k_\perp$. In Figure 7-5 we show $|P_b' - P_b|$ and see that at high redshift and cool heating models, the forest power spectrum is mostly uniform in $\perp$ though at lower redshifts and hotter IGM, there is significant $k_\perp$ structure. Since in section 7.2 we showed that P_f only varies along $k_\parallel$, this suggests that the cross power spectrum, $P_{f,b}$ is the prime contributor to $P_b' - P_b$ in a hot IGM, while P_f is in a cool one. The trough at lower redshifts, at $k \sim 0.5$ Mpc^{-1} is caused by the fact that $-2P_{f,b}$ is negative as we shall see below.

A potentially interesting consequence of the auto-terms invariance in $k_\perp$ is a potential for contaminating the separation of powers analysis advocated in Barkana and Loeb [13]. We may Taylor expand P_f

$$P_f(k_\parallel) = P_f(k\mu) = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial P_f}{\partial(\mu k)}\bigg|_{\mu k=0} (\mu k)^n, \tag{7.22}$$

so P_f introduces signal over a wide range of powers of μ and has the potential to contaminate the cosmological μ^4 and μ^6 components of the brightness temperature power spectrum. On the other hand, the small k, where the perturbative expansion is most accurate, is dominated by the diffuse brightness temperature emission. In all but the coolest heating models, contamination will likely be small, since we can see in Figure 7-3 that $P_f \lesssim 0.1 P_b$ at $k \lesssim 0.1$ Mpc^{-1}.

Decomposing the forest signal into powers of μ may be another way of distinguishing it from the brightness temperature. Even within the "IGM dominated" region. Detailed analysis on contamination of the cosmological signal and additional distinguishability offered by the angular dependence is beyond the scope of this paper will be the subject of future work.

To be more quantitative, we turn our attention to right hand side of Equation (7.15) and verify our decomposition of the forest power spectrum into $P_{\tau_{21}}^{LoS}$ and the sum of background source fluxes. To do this, we find the summed squares of the fluxes (at the center frequency of the observation) of all sources falling in or behind our data cubes at several redshifts, multiply by the 1D LoS power spectrum of τ_{21}

and compare with Δ_f^2 computed from our simulation as outlined above. We find that Equation (7.15) consistently underpredicts the simulation amplitude by a factor of 2. However, when we remove the clustering of sources by randomly assigning source positions (rather than using the dark matter biased positions), Equation (7.15) agrees with simulation output within $5 - 20\%$ over the studied redshifts. Hence we rewrite Equation (7.16) as

$$P_f \approx A_{cl}\frac{cD_M^2\lambda^4}{4k_B^2}P_{\tau_{21}}^{LoS}\int_z^\infty\int_0^\infty s^2\rho(z,z',s)\frac{D_M^2(z')}{H(z')}dz'ds \qquad (7.23)$$

Where A_{cl} is a constant of order unity that accounts for the boost in power due to clustering. We briefly explain this power boost in Appendix 7.A. In Figure $7-6$ we show the power spectrum, Δ_f^2 computed from our simulation and the prediction from Equation (7.15) for several redshifts in our fiducial heating model. For $k \gtrsim 10^{-1}$ Mpc^{-1}, Equation (7.15) agrees with our simulation at the 10% level, indicating that we can ignore the cross terms in Equation (7.14) and consider the forest power spectrum as the simple product of the 1D τ_{21} power spectrum and the integrated radio luminosity function.

A striking feature of Figure 7-5 is the apparent similarity of P_f along diagonal sets of different redshifts and models. For example, the "Cool IGM" model at $z = 12.2$ is very similar to the "Fiducial IGM" result at $z = 15.4$ and the "Hot IGM" at $z = 17.5$. It is suggestive that one can obtain the results of one particular thermal model by simply shifting another model in redshift, this translational invariance in redshift demonstrates that we may not need to simulate a broad range of heating models to understand the evolution of the forest power spectrum. Indeed, given our decomposition in Equation (7.15) where the amplitude of P_f is proportional to $\langle\tau_{21}\rangle^2 \propto \langle T_s^{-1}\rangle^2$, we should expect $\langle T_s\rangle$ to be a more generally applicable parameterization than f_X and redshift during the pre-reionization epoch. To show the importance of $\langle T_s\rangle$ as a parameter, we plot, in Figure 7-7, the amplitude of P_f at $k_\parallel = 0.5$ Mpc^{-1} as a function of $\langle T_s\rangle$ for our three heating scenarios and redshifts. Across all thermal models and redshifts, the amplitude of P_f is well described by a power law of $\langle T_s\rangle^{-2}$,
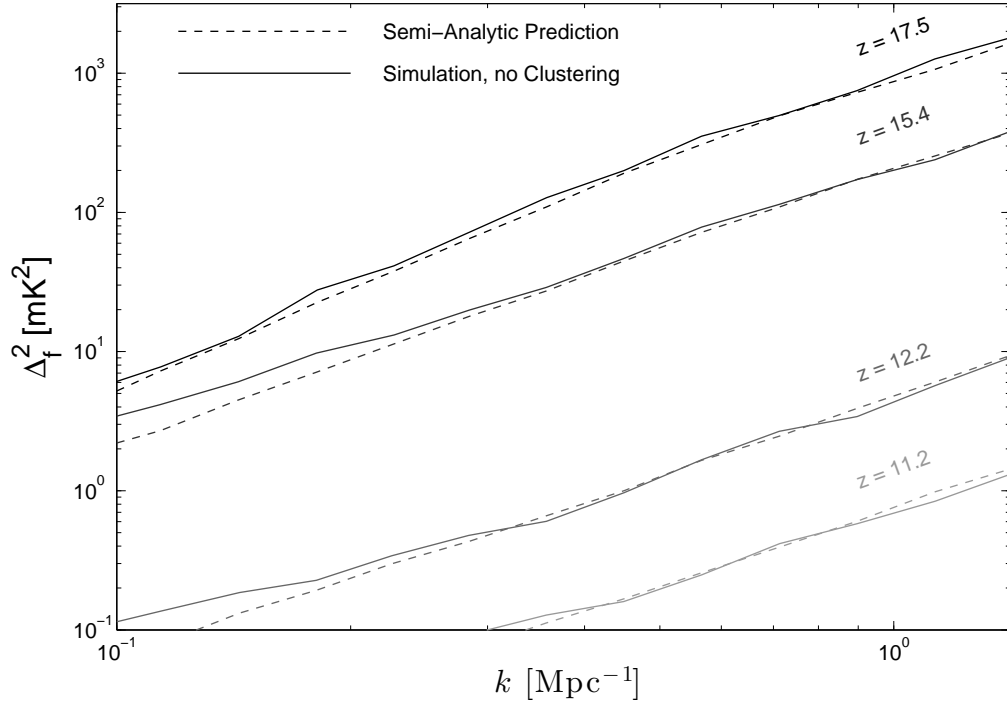
Figure 7-6: Our semi-analytic prediction agrees well with unclustered simulation results. The semi-analytic prediction of Equation (7.15) is plotted with dashed lines and $\Delta_f^2(k)$ computed directly from our simulation without clustering in solid lines. This demonstrates that for $k \gtrsim 10^{-1}$ Mpc^{-1}, the cross terms in Equation (7.14) may be ignored and P_f may be well approximated by the LoS power spectrum of τ_{21} multiplied by the summed squared fluxes for sources lying in and behind the data cube.

consistent with the normalization predicted in Equation (7.15).

Verifying our prediction on the sign of $P_{f,b}$ is our next task; we plot this quantity in Figure 7-8 for all models and redshifts. At high redshift, $P_{f,b}$ is entirely negative due to the anti-correlation between T_f and T_b and adds to the total amplitude of P_b'. As heating takes place, T_s drops out of T_b and fluctuations in T_b are sourced predominantly by variations in x_{HI} leading to positive correlation between T_b and T_f for positive $P_{f,b}$. As we see from the figures, this process is "inside-out", with large scales remaining anti-correlated longer than the small scales. Heating proceeds in an "inside-out" manner, and since there is an overlap between the completion of heating and onset of reionization, temperature fluctuations remain important on large scales [187, 146].

7.5 Prospects for Detection with an MWA-like Array

We now turn to addressing the detectability of the power spectrum signature of the forest and its distinguishability from the power spectrum, P_b. Our strategy is to combine our simulations with random realizations of instrumental noise and galactic and extragalactic foregrounds. With data cubes containing both our simulated signals and our random contaminants, we can then take advantage of the full quadratic estimator formalism developed by Tegmark [208], adapted for 21 cm tomography by Liu and Tegmark [120]; hereafter LT11, and accelerated for large data sets by Dillon et al. [58]; hereafter D13. In this section, we will explain those techniques and show what results when our simulations of the forest are added to realistic foregrounds and instrumental noise.

7.5.1 Power Spectrum Estimation Methods

To estimate the power spectrum of the forest, we apply the quadratic estimator formalism [208]. This formalism has the advantage that, in the approximation of
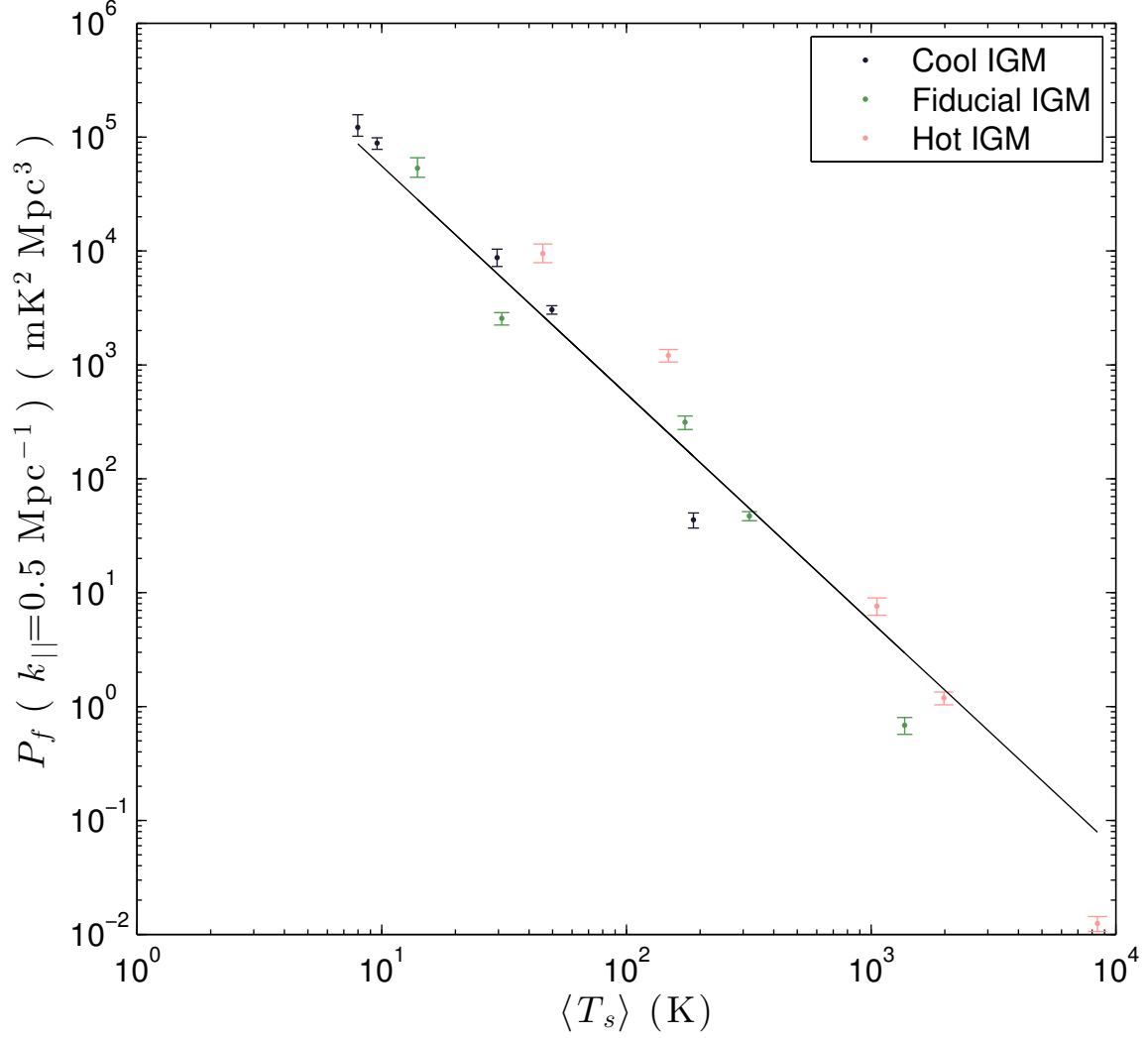
Figure 7-7: We see that for a fixed quasar distribution, the magnitude of P_f can be parameterized by $\langle T_s \rangle$ and that the amplitude is consistant with a simple power law. Here, we plot $P_f(k_\parallel)$ at $k_\parallel = 0.5$ Mpc^{-1} vs. $\langle T_s \rangle$ for all considered redshifts and f_X. The black line is the power law $\langle T_s \rangle^{-2}$ as one might expect for an amplitude set by $\langle \tau_{21} \rangle^2$ (Equation (7.16)). Inasmuch of this simple trend, a modest spread in heating models gives us a decent understanding of the behavior of the amplitude for P_f. This relation holds for the quasar population considered here because the integral over the luminosity function does not change significantly over the redshifts we consider.
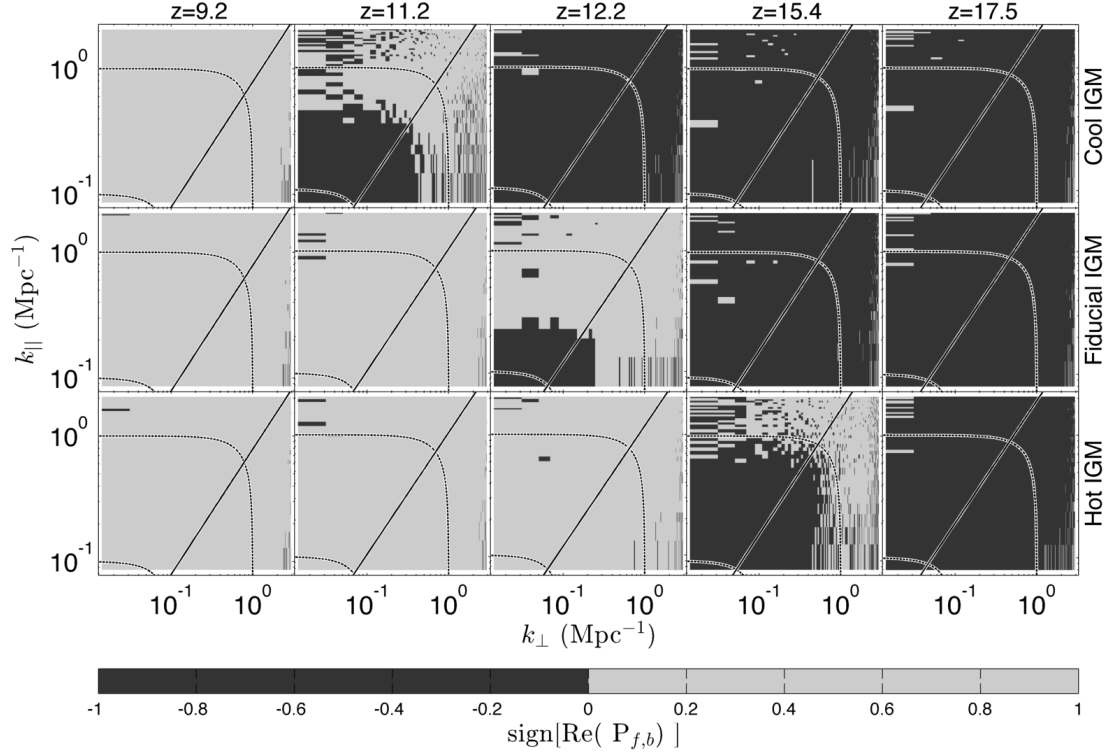
Figure 7-8: The cross power spectrum, $\mathrm{Re}(P_{f,b})$'s, sign is determined by the anti-correlation of x_{HI} and T_s during the pre-heating epoch and by x_{HI} after heating has taken place. Here we show the sign of $\mathrm{Re}(P_{f,b})$ for our three different heating models as a function of redshift. At pre-heating redshifts, T_s is small and x_{HI} is relatively uniform so that T_b and T_f primarily depend on T_s and anti-correlate so that $\mathrm{Re}(P_{f,b})$ is negative. At low redshifts, T_b is independent of T_s and fluctuations are primarily sourced by x_{HI} so that T_b and T_f are correlated and $\mathrm{Re}(P_{f,b})$ is positive. Futhermore, heating proceeds in an "inside-out" manner so that the smallest scales become correlated first.

foregrounds and noise that are completely described by their covariances, all cosmological information is preserved in going from three-dimensional data cubes to power spectra. This formalism was adapted by LT11 for 21 cm power spectrum estimation and further refined and accelerated by D13.[10]

In essence, the method relies an optimal and unbiased estimator of band powers in the $k_\perp$-$k_\parallel$ plane, $\widehat{\mathbf{p}}$, defined as

$$\widehat{p}^\alpha = \sum_\beta M^{\alpha\beta} \left(\mathbf{x}^\mathsf{T} \mathbf{C}^{-1} \mathbf{Q}^\beta \mathbf{C}^{-1} \mathbf{x} - b^\beta \right). \tag{7.24}$$

where $\mathbf{x}$ is a vector containing mean-subtracted data, $\mathbf{C}$ is the covariance of $\mathbf{x}$, including noise and contaminants, $\mathbf{Q}$ is a matrix that encodes the Fourier transforming, squaring, and binning necessary to calculate a band power, and $\mathbf{b}$ is the bias term. The normalization matrix $\mathbf{M}$ is related to the Fisher information matrix $\mathbf{F}$. Both $\mathbf{F}$ and $\mathbf{b}$ can be calculated via a Monte Carlo using the fact that

$$b^\beta = \langle \mathbf{x}^\mathsf{T} \mathbf{C}^{-1} \mathbf{Q}^\beta \mathbf{C}^{-1} \mathbf{x} \rangle \equiv \langle \widehat{q}^\beta \rangle \tag{7.25}$$

and that

$$\mathbf{F} = \mathrm{Cov}(\widehat{\mathbf{q}}). \tag{7.26}$$

The ensemble average of each band power is related to the true band power $\mathbf{p}$ by a window function matrix, $\mathbf{W} = \mathbf{MF}$,

$$\langle \widehat{\mathbf{p}} \rangle = \mathbf{W}\mathbf{p}. \tag{7.27}$$

The error on true band powers is also related to $\mathbf{M}$ and $\mathbf{F}$ through

$$\mathrm{Cov}(\widehat{\mathbf{p}}) = \mathbf{M}\mathbf{F}\mathbf{M}^\mathsf{T}. \tag{7.28}$$

Each quadratic estimator can thus be thought of as a weighted average of the true

---

[10]For further details on this particular implementation of the quadratic estimator method, the reader is referred to D13.

band powers with potentially correlated errors, both of which depend on one's choice of **M**. Though any choice of **M** that makes **W** a properly normalized weighted average is reasonable, we adopt a form of **M** that makes the errors on $\widehat{\mathbf{p}}$ uncorrelated. Dillon et al. [59], argue that this choice of **M** dramatically reduces the contamination of the EoR window by residual foregrounds. It also provides a set of band power estimates which can be considered both mutually exclusive and collectively exhaustive because they cover the whole $k_\perp$-$k_\parallel$ plane while not containing any overlapping information.

7.5.2 Noise and Foreground Models

The method outlined above requires model means and covariances of the contaminants that contribute to **x**, like noise and foregrounds. Our model of the instrumental noise depends, first and foremost, on the design of the interferometer. In this paper, we consider the MWA with 128 tiles whose locations are detailed in Beardsley et al. [16] as representative of the current generation of low frequency interferometers. Additionally, we consider possible realizations of double and quadruple sized instruments (MWA-256T and MWA-512T, respectively), as representative of extensions to current generation interferometers or next generation, $A_{eff} \sim 0.1\text{km}^2$, arrays such as the Hydrogen Epoch of Reionization Array (HERA) [7]. As we will show, we generally do not need a square kilometer scale instrument to see the statistical effects of the forest.

To generate our MWA-256T and MWA-512T designs with maximum sensitivity to 21 cm cosmology, we add antenna tiles to the current MWA-128T design within a dense core 900 m in radius. These are drawn blindly from a probability distribution similar to that in Bowman et al. [26]: uniform for $r < 50$ m and decreasing as r^{-2} above 50 m. The tile locations of the arrays we use are shown in Figure 7-9.

Our model for the noise is adapted from D13 [11]. In it, we incorporate observation times calculated in each uv-cell from 1000 hours of rotation synthesis at the latitude of the MWA. The effective area of each tile is computed using a crossed dipole

---

[11]The method of D13 is adapted with one correction: the form of the noise power spectrum adapted from [210] does not include the assumption that the field and beam sizes are the same.
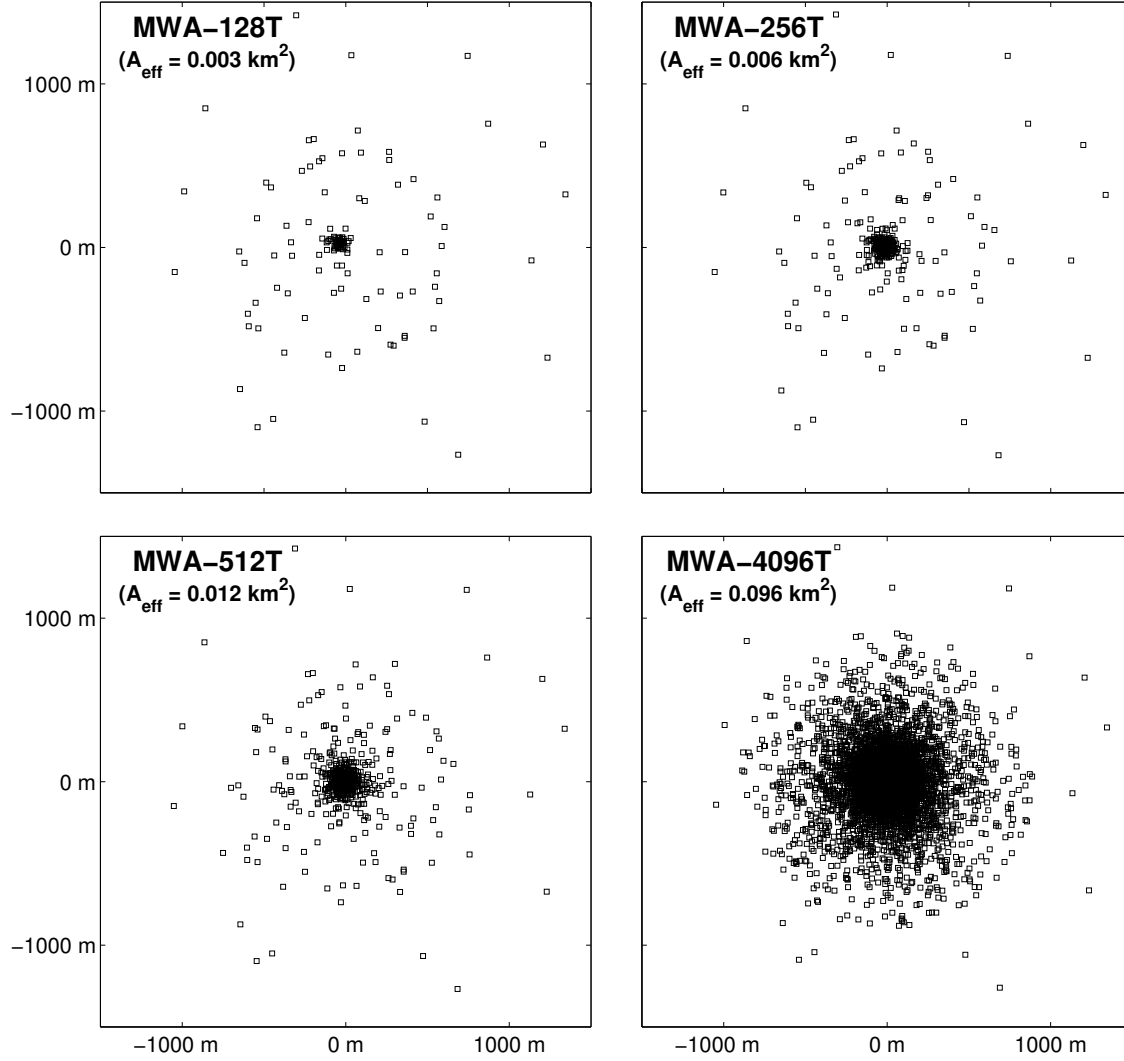
Figure 7-9: Array layouts that we use to determine the detectibility and distinguishability of the 21 cm forest power spectrum signature. We chose to study two moderate extensions of the MWA-128T: MWA-256T and MWA-512T. In addition we study a 4096T array that is representative of a HERA scale instrument with ~ 400 times the collecting area of the MWA. Tile locations are drawn randomly from a distribution that is constant for the inner 50 m and drops as r^{-2} for larger radii.

f (MHz)	FWHM (deg)	A_{eff} (m^2)	T_{sys} (K)
150	23	23	290
120	30	24	490
100	34	24	760
80	39	27	1300

Table 7.2: Instrumental Parameters

model while the system temperature is treated as the sum of receiver temperature, given by a power law fitted to two data points appearing in Tingay et al. [220], and sky temperature, measured in Rogers and Bowman [191]. In Table 7.2 we give our instrumental parameters at several different frequencies.

Similarly, our model of the foregrounds is the one application of the model developed by LT11 and D13. For the sake of simplicity,[12] we model extragalactic foregrounds as a random field of point sources with fluxes up to 200 Jy. They have an average spectral index of 0.5 and variance in their spectral indices of 0.5. Their clustering has a correlation length scale of 7'. Likewise, we model Galactic synchrotron radiation as a random field with an amplitude of 335.4 K at 150 MHz, a coherence length scale of 30°, and a mean spectral index of 0.8 with an uncertainty in that index of 0.1.

As we have previously discussed, we conservatively cut out the region of $k_{\perp}$-$k_{\parallel}$ space that lies below the wedge. Once the wedge has been excised, we optimally bin from 2D to 1D Fourier space with the inverse covariance weighted technique described by D13.

To create simulated observations, we divide our simulated volumes into 36 fields, each 750 Mpc on a side, which roughly fill the primary beam of our antenna tiles. We add random noise and foregrounds to each field independently, taking advantage of the fast technique for foreground and noise simulations developed by D13. Finally, we take the sample variance of the cosmic signal into account by using our power spectrum results from Section 8.3 and by counting the number of independent modes

---

[12]Breaking extragalactic foregrounds into a bright "resolved" population and a confusion-limited "unresolved" population only improves the error bars (D13), so our efficient choice is also a conservative one.

probed by the instrument at each k scale.

7.5.3 Detectability Results

We now present the results of our sensitivity calculation. We demonstrate that, given prior knowledge[13] of the X-ray heating history, a power spectrum measurement with a modest expansion of an MWA-like instrument is sufficient to distinguish between scenarios with or without the forest in our fiducial and cool heating models. Since the forest signal is detectable with smaller arrays only at smaller k, where P_b dominates, its effect is likely degenerate with diffuse IGM emission. Observing this region for all considered models will require a HERA scale instrument with $A_{eff} \sim 0.1\text{km}^2$.

In order to determine the array size necessary to resolve the forest power spectrum, we first focus on $z = 11.2$, the lowest redshift considered where there is significant signal for one of our thermal models and quasar counts are relatively high. In Figure 7-10 we shade the 2σ region for a detection of $\Delta^2(k)$ with no 21 cm forest absorption present and mark detections of $\Delta^2(k)$ with 21 cm forest absorption with black dots. The 2σ vertical error bars, given by the diagonal elements of Equation (7.28) are marked in red. Also marked in red are the horizontal error bars which are given by the 20th and 80th percentiles of the window functions. To determine whether we can detect the forest imprint, we ask "are the points consistent with the gray shaded region?"

We see that MWA-256T and MWA-512T can distinguish cool models with and without the forest at greater than 2σ. However these detections are not within the region of Fourier space where the forest dominates P_b'. As a result, though MWA expansions can resolve two models with or without the forest, it is unlikely that they will be able to distinguish a model with the forest from one with a slight variation in heating. If an independent measure of the global spin temperature can be obtained, the radio luminosity function might be constrained with a modest MWA extension. We note that MWA-4096T is only able to detect the forest in our cool model at

---

[13]Here, "prior knowledge" means that we know what the IGM power spectrum without the 21 cm forest to within the error bars of our thermal senstivity.

z = 11.2

Cool IGM   Fiducial IGM   Hot IGM

Δ^2 (mK^2) with MWA−128T (A_{eff} = 0.003 km^2)

Δ^2 (mK^2) with MWA−256T (A_{eff} = 0.006 km^2)

Δ^2 (mK^2) with MWA−512T (A_{eff} = 0.012 km^2)

Δ^2 (mK^2) with MWA−4096T (A_{eff} = 0.096 km^2)

k (Mpc^{-1})

·   Detections with RL Sources   ▼   2σ Limits with RL Sources   Detections or 2σ Limits without RL Sources
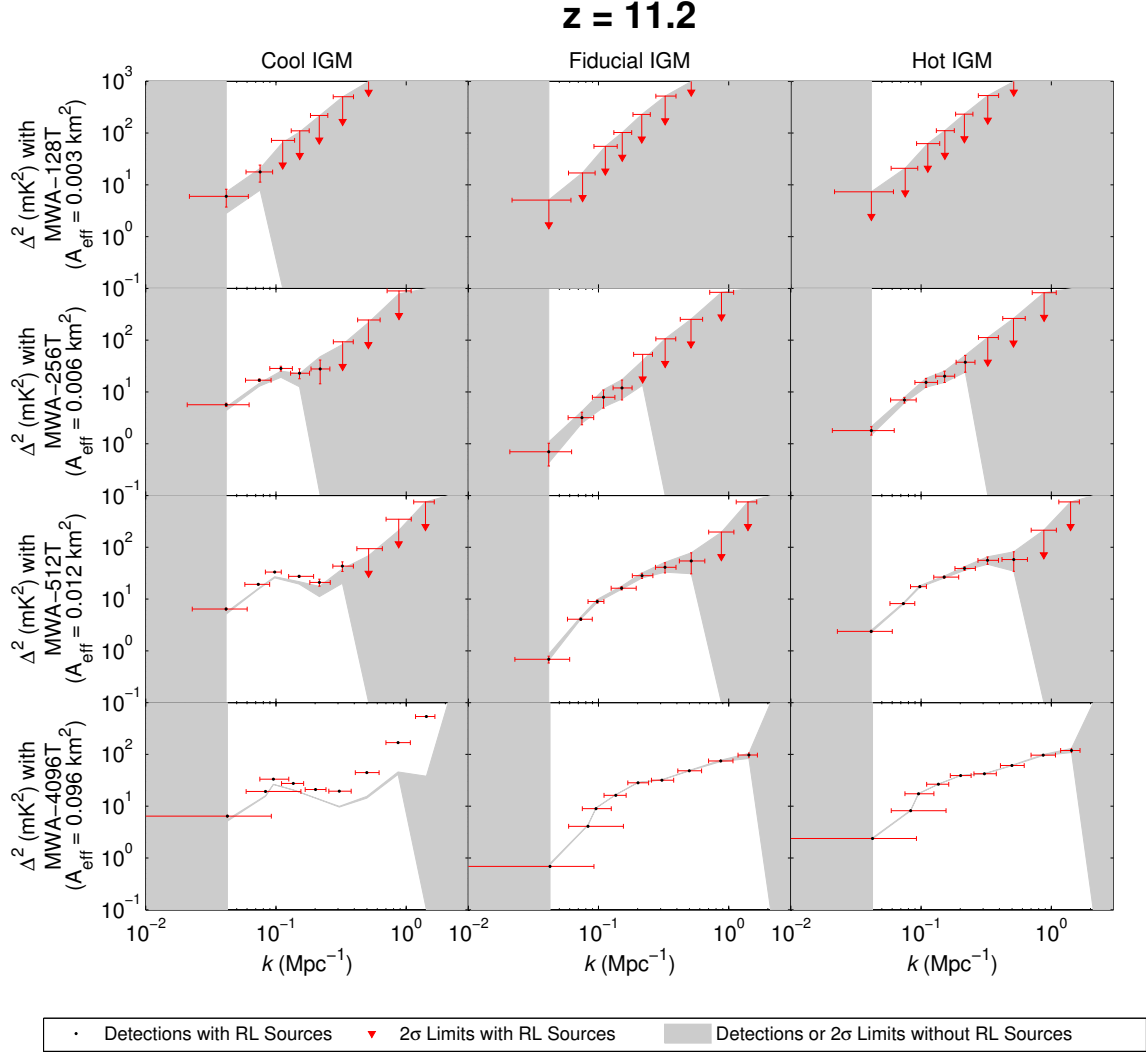
Figure 7-10: Detections (black dots) and upper limits (red triangles) of the 21 cm Power Spectrum at z=11.2 for all of our arrays and heating models in the presence of 21 cm forest absorption from background RL sources. The grey fill denotes the $2\ \sigma$ region around the measured power spectrum with no RL sources present. To determine whether we can detect the forest imprint we ask, "do the points and their error bars lie outside the gray shaded region?" MWA-256T and MWA-512T would be capable of distinguishing power spectra with or without sources in our cool IGM model, however only 4096T is consistantly sensitive to the $k \gtrsim 0.5$ Mpc^{-1} region where the forest dominates. Only for our cool IGM model, MWA-512T would suffi-cient to detect this upturn as well. Hence a moderate MWA extension would likely be able to constrain some RL populations given a cooler heating scenario while a HERA scale instrument will be able to constrain the W08 RL population using the Forest power spectrum even for more emissive heating scenarios. Note that the upturn in the gray region is not from increased power at high k but larger error bars.

$z = 11.2$ since the optical depth in our more X-ray emissive models is far too small at this time.

To see more broadly what might be achieved by the next generation, we show in Figure 7-11 the error bars and detections with and without the forest across all considered f_X and z for our HERA scale model. We find that z=15.4 is our "sweet spot" for the W08 distribution. 4096T is able to resolve the $k \gtrsim 0.5$ Mpc^{-1} forest region for all of the IGM heating models that we investigate. For our cool and fiducial models, 4096T is also able to observe the forest region for a range of redshifts. These results show that a HERA scale array has the potential to constrain the IGM state by measuring Δ^2 for $k \lesssim 10^{-1}$ Mpc^{-1}, where the brightness temperature dominates, and the RL distribution in observing the region $k \gtrsim 0.5$ Mpc^{-1} where the forest has a significant contribution.

Over the course of the IGM's evolution, there are times where the 21-cm power spectrum becomes particularly steep; for example, during the era immediately before the X-ray heating peak. As a result, observing excess power at $k \gtrsim 0.5$ Mpc^{-1} for a single redshift alone will likely not be sufficient to constrain the radio luminosity function. However, discerning the IGM thermal history with measurements of the power spectrum amplitude at $k \sim 0.1$ Mpc^{-1} and observing an absence of flattening at high k, over the range of redshifts after the X-ray heating peak as shown in Figure 7-4 should allow for constraints to be placed on the high-redshift radio luminosity function.

7.5.4 Distinguishability Results

In order to quantify how distinguishable our simulations with the forest are from our simulations without the forest for a given instrument, redshift, and heating model, we calculate the standard score of the χ^2 sum of the power spectrum values across all k-bins,

$$Z \equiv \frac{\chi^2 - N_k}{\sqrt{2N_k}},$$
(7.29)
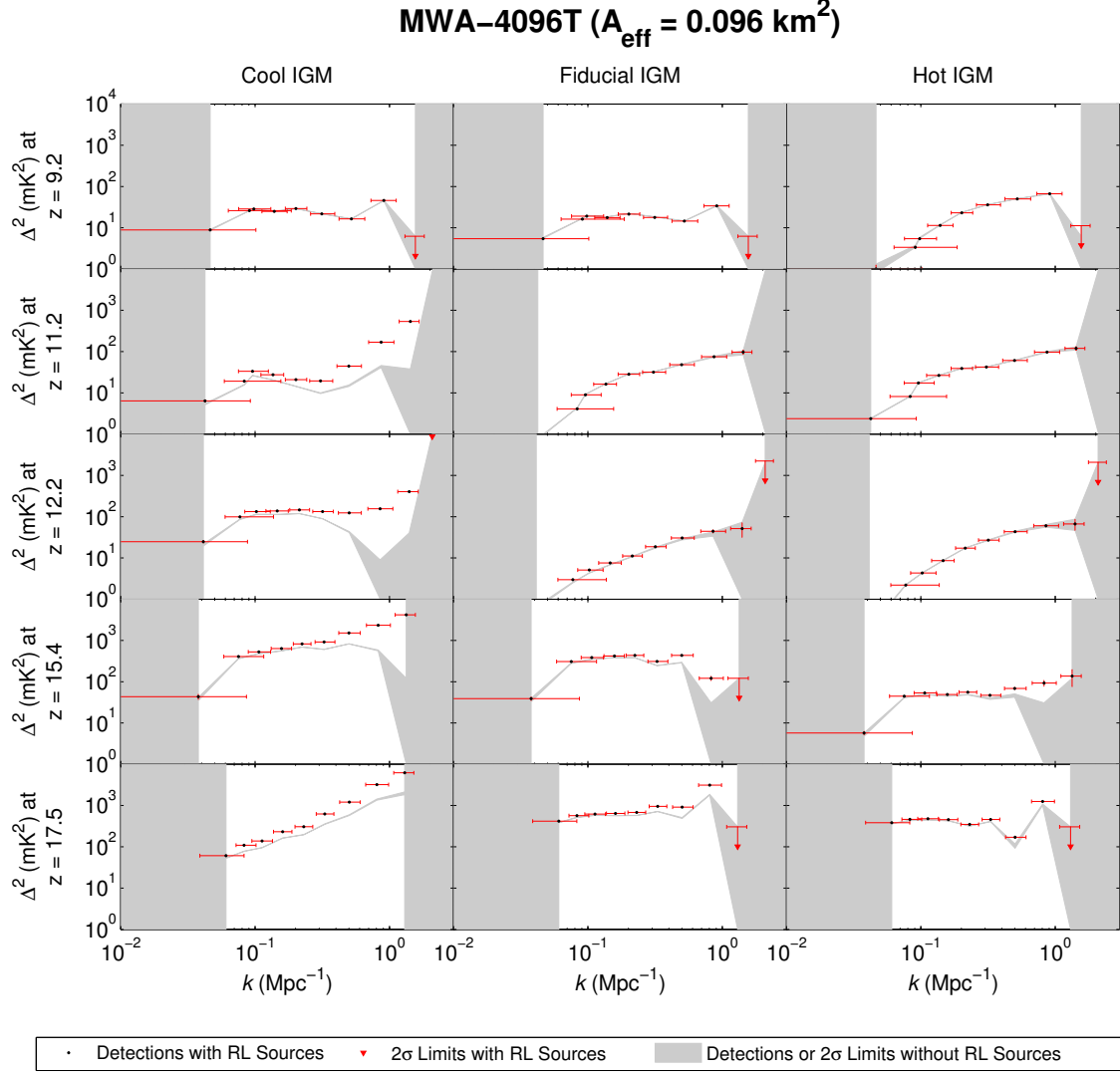
Figure 7-11: These plots are identical to Figure 7-10 except the array is fixed to be MWA-4096T, representative of a HERA generation instrument, and redshift is varied. A HERA class instrument is able to resolve the upturn at $k \gtrsim 0.5$ Mpc^{-1} that distinguishes the forest, and should be able to detect the 21 cm forest feature considered in this work for a variety of heating scenarios. The thermal noise error bars are to small to resolve by eye in most of these plots.

where N_k is the number of k bins, $\chi^2 \equiv \sum_k \left(\frac{P_b(k)' - P_b(k)}{\sigma_k} \right)^2$, and σ_k are the diagonal elements of Equation (7.28) for each model without the 21 cm forest present. Assuming statistical independence between k bins, Z is the number of standard deviations at which we can distinguish a model with the 21 cm forest from a model without it using the χ^2 statistic. Unfortunately, this measure is somewhat naive since it does not account for potential degeneracies in the power spectrum amplitude from different thermal histories. However it enables us to quantitively compare outlooks across the numerous dimensions of redshift, array, and heating history. We consider a $Z \gtrsim 10$ to indicate significant distinguishability.

In Figure 7-12 we show the value of Equation (7.29) for all models and arrays. Our first observation is that MWA-128T is not capable of distinguishing a model with the forest from a model without the forest for any of the considered f_X. MWA-256T would be capable of distinguishing the forest at all considered $z \gtrsim 9.2$ for our cool X-ray heating model at greater than 5σ and in our fiducial heating model only at the highest considered redshift (which is near the X-ray heating peak). MWA-512T would be capable of resolving the forest at the two highest redshifts for our fiducial model and at all considered redshifts for our cool model. The hot model remains unobservable for all MWA expansion arrays but is accessible to a HERA scale instrument.

How the distinguishability between different heating models is affected by the presence of the 21 cm forest is explored in Figure 7-13. In our 128T table, we see that a detection of the IGM and constraints on low X-ray emissive histories are possible with the current generation of EoR experiments. There are several caveats worth noting however. First, the high S/N distinctions at $z = 9.2$ are due to a detection of the reionization peak at redshifts in which reionization physics such as the uv-efficiency (which we have assumed fixed) become significant. However, we note that this result contradicts the marginal detectability claimed in Mesinger et al. [147] primarily due to the fact that we include bins with $k < 0.1\ \mathrm{Mpc}^{-1}$ in our standard score. Though these bins have large S/N they may be contaminated by more pessimistic foreground leakage than we consider here such as what is observed

by Pober et al. [182]. We also note that the increased sensitivity of combining k-bins allow for constraints on the fiducial X-ray model at $z \sim 15$. The peaks in detectability at $z \approx 9$ and $z \approx 17$ arise from the two peaked structure of the power spectrum in redshift with the low redshift peak corresponding the reionization, and the high redshift peak corresponding to x-ray heating [187]. We see that the forest introduces a small enhancement to the distinguishability between hot and cool heating models. Since the forest adds positively to the power spectrum of a cool, optically thick IGM, its presence enhances the distinguishability between vigorous and cool heating. We find that a modest extension to the MWA can distinguish between hot and fiducial models over a wider range of redshifts and MWA-4096T is able to distinguish between all models over our entire considered redshift range.

7.6   The Detectability of the Forest over a Broad Parameter Space

For the sake of simplicity, we focus on the detectability of the 21 cm Forest power spectrum from the single population model considered in Wilman et al. [236]. In doing this, it is unclear over what range of radio loud populations the signal is observable. Fortunately, thanks to Equation (7.23), we can give order of magnitude estimates of how the detectability of the Forest power spectrum scales with the radio loud source population and the heating history. According to Equation (7.23), the amplitude of the forest power spectrum, at prereioinization redshifts, scales as

$$P_f \propto \frac{1}{\langle T_s \rangle^2} \frac{\sum_i s_i^2(> z)}{\Omega} \tag{7.30}$$

where $\sum_i s_i^2(> z)\Omega^{-1}$ is the average sum of source fluxes squared per solid angle. We will call this quantity the *flux squared density* of the source population. We take advantage of the simple scaling in Equation (7.30) to extrapolate the amplitude of the Forest signal over a large range of heating models and redshifts. At each redshift, with our fiducial heating model and source population, we obtain a normalization factor

MWA–128T (A$_{eff}$ = 0.003 km^2)

	Cool IGM: With vs. Without Forest	Fiducial IGM: With vs. Without Forest	Hot IGM: With vs. Without Forest
z=9.2	-1.3σ	-0.9σ	0.77σ
11.2	-1σ	-1.2σ	-0.97σ
12.2	-0.31σ	-0.23σ	0.64σ
15.4	0.65σ	0.18σ	2.1σ
17.5	-1.3σ	1σ	0.8σ

MWA–256T (A$_{eff}$ = 0.006 km^2)

	Cool IGM: With vs. Without Forest	Fiducial IGM: With vs. Without Forest	Hot IGM: With vs. Without Forest
z=9.2	-1.3σ	-0.86σ	5.1σ
11.2	2.3σ	-1.3σ	-0.99σ
12.2	12σ	0.0057σ	0.67σ
15.4	12σ	0.63σ	1.7σ
17.5	-1.1σ	7.1σ	2.3σ

MWA–512T (A$_{eff}$ = 0.012 km^2)

	Cool IGM: With vs. Without Forest	Fiducial IGM: With vs. Without Forest	Hot IGM: With vs. Without Forest
z=9.2	-0.98σ	-0.77σ	18σ
11.2	50σ	-1.2σ	-1.1σ
12.2	50σ	0.92σ	0.71σ
15.4	54σ	12σ	1.8σ
17.5	6.1σ	24σ	6.8σ

MWA–4096T (A$_{eff}$ = 0.096 km^2)

	Cool IGM: With vs. Without Forest	Fiducial IGM: With vs. Without Forest	Hot IGM: With vs. Without Forest
z=9.2	24σ	-0.52σ	$1 \times 10^2\sigma$
11.2	$4.3 \times 10^3\sigma$	27σ	-0.035σ
12.2	$4.6 \times 10^3\sigma$	69σ	2.6σ
15.4	$1.2 \times 10^4\sigma$	$9.8 \times 10^2\sigma$	91σ
17.5	$2.7 \times 10^3\sigma$	$1.4 \times 10^3\sigma$	$1.2 \times 10^2\sigma$

$\log_{10}[\text{Z–Score of } \chi^2 \text{ with N}_k \text{ DoF}]$

0   0.5   1   1.5   2   2.5   3   3.5   4

Figure 7-12: The significance of distinguishability across all measured k bins (Equation (7.29)) for all arrays, redshifts, and IGM heating models for a 1000 hour observation. An extension of MWA-128T is capable of distinguishing models with and without the 21 cm forest from the W08 RL population in our cool and fiducial heating scenarios. MWA-512T and HERA scale MWA-4096T are capable of distinguishing the forest in the power spectrum in all heating models considered in this work.
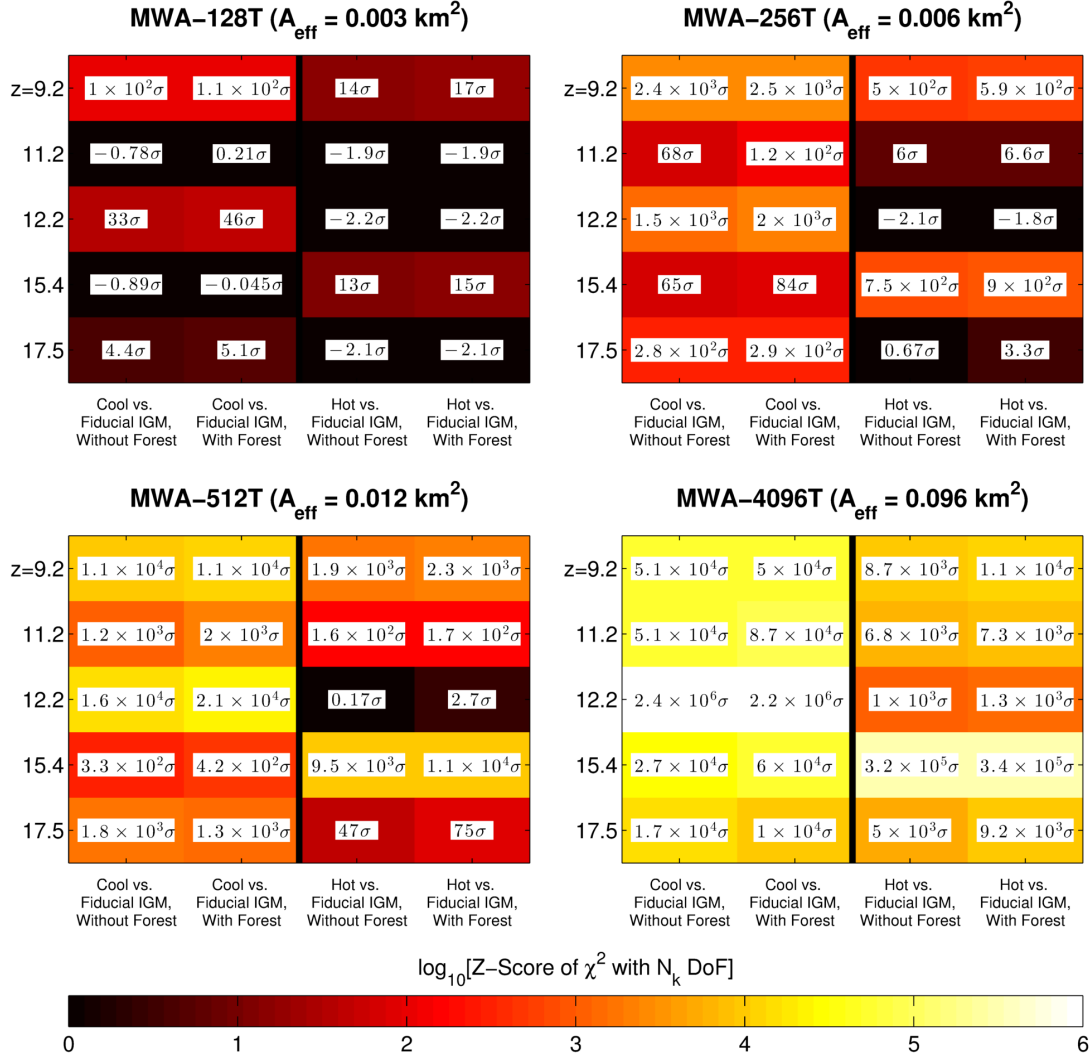
Figure 7-13: The 21 cm Forest moderately enchances the distinguishability between thermal scenarios and MWA scale interferometers can distinguish between the power spectra for reasonable X-ray heating histories. Here we show the cumulative z-score described in Equation (7.29), except now applied to the difference between different IGM heating models, for all arrays and redshifts. At low redshift, the forest decreases the distinguishability of different X-ray heating scenarios by subtracting from the higher amplitude model. When the positive auto-term dominates at high redshift, the forest increases the contrast between given heating models.

for P_f at a single mode, $k = 0.5$ Mpc^{-1}. We then compute $\langle T_s \rangle$ for a large number of lower resolution, $(600 \text{ Mpc})^3$ 21cmFAST simulations with 400^3 pixels, varying the f_X parameter by three orders of magnitude from $f_X = 10^{-2} - 10^1$. In Figure 7-14, we show the ratio of P_f to the amplitude of thermal noise as a function of f_X and the flux squared density of sources, marking the predicted flux squared density of Wilman et al. [236] by a dashed black and white line. We find that the detectability of the forest power spectrum at $z \sim 10$ depends strongly on the thermal state of the IGM, with models significantly fainter than Wilman et al. [236] undetectable except for cool heating histories with $f_X \lesssim 10^{-1}$. On the other hand, for $z \gtrsim 15$, X-rays in all models have not had sufficient time to heat the IGM above the adiabatic cooling floor and the detectability of P_f becomes significantly less dependent on f_X, allowing for a broader range of populations to be probed at higher f_X.

7.7 Conclusions and Future Outlook

Using semi-numerical simulations of the thermal history of the IGM, and a semi-empirical RL source distribution, we have shown that the 21 cm forest imprints a distinctive feature in the power spectrum that is, for the most part, invariant in $k_\perp$ and, depending on the RL population and thermal history, potentially dominates over the cosmological 21 cm power spectrum at $k_\parallel \gtrsim 0.5$ Mpc^{-1}. We have also derived a simple semi-analytic equation that directly relates the forest power spectrum of τ_{21} and the radio luminosity function.

Using realistic simulations of power spectrum estimation and including the effects of foregrounds and noise, we have shown that a moderate extension of the MWA-128T instrument has the thermal sensitivity to detect the forest feature in the power spectrum for the W08 RL population with an X-ray efficiency of $f_x \lesssim 1$. For more vigorous heating scenarios, a HERA scale array will have the sensitivity to distinguish this feature. Our simulations also support the results of Christian and Loeb [42] and Mesinger et al. [147], that low emissivity heating scenarios can be constrained with existing arrays and an extensive examination of the heating history will be possible
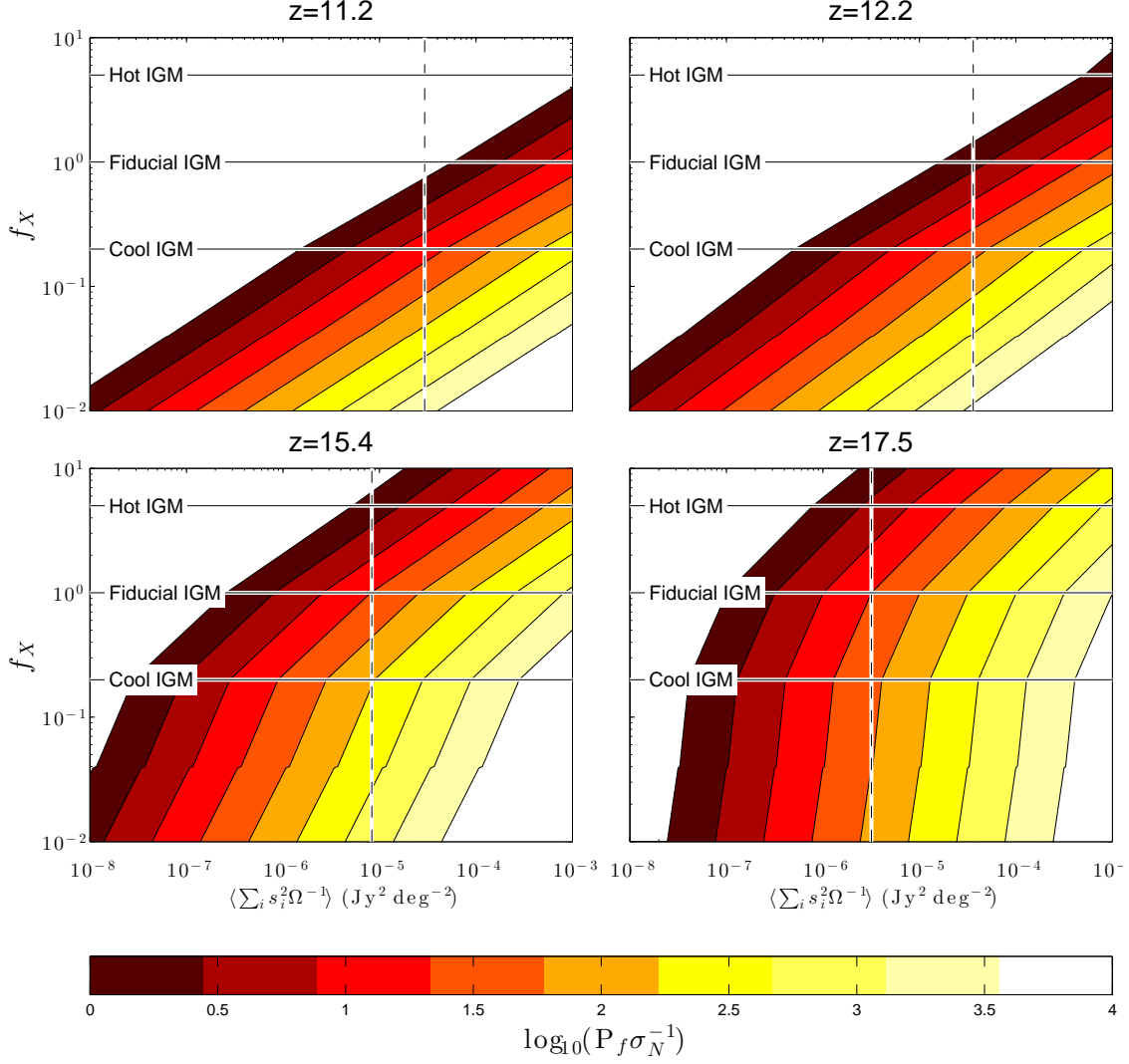
Figure 7-14: The ratio of the 21 cm Forest power spectrum, $P_f(k = 0.5 \text{ Mpc}^{-1})$ to thermal noise for 1000 hours of observation on a HERA scale interometer, extrapolated over a large range of X-ray efficiencies and flux squared densities. Vertical dashed black and white lines indicate the value of the simulation by [236] while the horizontal black and white lines indicate the f_X efficiencies that we explicitly simulate in this paper. At the highest redshifts, $\langle T_s \rangle$ levels off and the detectability of the signal is independent of redshift. At late prereionizatoin redshifts, we see that the 21 cm Forest will only be detectable for heating efficiencies $\lesssim 1$.

in the future with larger instruments.

Signal-to-noise considerations alone do not tell us whether we will be able to distinguish the forest signal from the effects of IGM physics on the power spectrum, especially at small k where a slight change in f_X might shift the power spectrum amplitude up or down, mimicking the shift from the 21 cm forest. Fortunately, the region, $k \gtrsim 0.5$ Mpc^{-1} is dominated by the forest power spectrum, P_f, for a range of redshifts in all of our heating models. Specifically, the 21 cm forest removes the $k \gtrsim 0.5$ Mpc^{-1} flattening that occurs after the X-ray heating peak. Observations of the power spectrum over a range of redshifts, with a sensitivity similar to HERA or the SKA should be able to isolate the thermal history at $k \lesssim 0.1$ Mpc^{-1} and constrain RL populations similar to that of W08 at $k \gtrsim 0.5$ Mpc^{-1}.

While this paper is a proof of concept, considering a single fiducial RL source distribution, it is possible that measurements with current generation instruments, or moderate extensions, can put constraints on more optimistic scenarios. On the other hand, there are many steep decline scenarios whose power spectrum signatures will be inaccessible even to future arrays. In section 7.6 we illustrate the scaling of the detectability of the signal with source flux squared density and X-ray emissivity, finding that populations with order of magnitude smaller flux squared densities than W08 will require a relatively cool prereionization IGM to be detectable. In particular, we note that the H04 simulation is one to two orders of magnitude more pessimistic than the predictions of W08 at the highest considered redshifts and would not be detectable in the forest dominated region if $f_X \gtrsim 10^{-1}$. However, higher resolution simulations of the IGM indicate that $\Delta^2_{\tau_{21}}$ continues to climb to $k \sim 10$ Mpc^{-1} while P_b remains flat. Hence the result of a fainter radio luminosity function would be to shift the region of forest dominance to higher k rather than eliminating it, leaving the possibility of detection for a more powerful instrument such as the SKA. There also exists the possibility of separating P_f using its LoS symmetry which might be exploited at $k \sim 0.1$ Mpc^{-1} where EoR interferometers are most sensitive. Finally, we have not considered the absorption of mini halos which Mack and Wyithe [132] show to substantially increase the variance along the line of sight towards sources

(see their Figure 11). Since this variance is an integral of the power spectrum we are being conservative in neglecting them. The sensitivity of future instruments to the forest can be enhanced by increased frequency resolution, allowing them to probe the higher $k_\parallel$ modes where the forest is especially strong. The parameter space of radio loud quasars is greatly unconstrained and the disparity between W08 and H04 simply underscores the need for future studies to explore this parameter space. The exploration of a range of RL populations for fixed arrays is left for future work.

In summary, we have shown that the 21 cm power spectrum not only contains information on the IGM in absorption and emission against the CMB but also includes detectible, and in many cases non-negligible signatures of the 21 cm forest. This absorption may be used to constrain the high redshift RL population and IGM thermal history with upcoming interferometers.

7.A   Appendix: A Derivation of the Morphology of P_f

In Section 7.2 we present a formula, Equation (7.15), for the the 21 cm forest power spectrum that is the sum of the auto power spectra along the line of sight to each background source. This equation is particularly convenient because it can easily be decomposed into an integral of the radio luminosity function and the optical depth power spectrum. In addition, its k-space morphology, which includes no structure in $k_\perp$, is relatively simple. In this appendix we derive Equation (7.15) by applying an analytic toy model to the auto and cross power spectrum contributions to P_f described in Equation (7.13). For the sake of analytic tractability, we invoke a number of approximations. However our results describe P_f very well for $k \gtrsim 10^{-1}$ Mpc^{-1}. Our assumptions are

1. The sources all have the same flux. The W08 simulation includes sources ranging from 1 nJy to ~ 10 mJy over the redshifts of interest. We see in Figure 7-2 that the integral of the source fluxes squared is dominated (at the 10% level) by

sources with S_ν between $1 - 10$ mJy so modeling our population as having equal flux gives a decent order of magnitude approximation.

2. The sources are spatially uncorrelated. Clustering from the W08 dark matter bias is actually significant and boosts the results of our simulation, relative to Equation (7.15), by a factor of two without changing P_f's predicted shape. We will thus absorb this clustering boost into a multiplicative factor of order unity.

3. The sources are unresolved. This will almost certainly be true in all interesting cases given the large synthesized beams of radio interferometers and the extreme distances to the sources.

4. Source spectra are flat over the frequency interval of a data cube. This is true on the 10% level over a ~ 8 MHz band for $S \sim \nu^{-.75}$ sources. Because this slow variation gives a very narrowly peaked convolution kernel in k-space, power spectra are not noticably effected by this assumption.

5. The source positions are completely uncorrelated with the cube optical depth field. In reality, the sources that fall within a data cube should be correlated with τ_{21}. We find that correlating or not correlating in cube sources only changes the simulation output by approximately 10%.

We start by reiterating Equation (7.14) where P_f may be written as

$$P_f = \frac{1}{V}\left\langle \left|\widetilde{\Delta T_{RL}\tau_{21}}\right|^2 \right\rangle = \sum_j P_j + 2\mathrm{Re}\left(\sum_{j<k} P_{j,k}\right) \equiv \Sigma_{auto} + \Sigma_{cross}, \qquad (7.31)$$

where $P_j = \frac{1}{V}\langle|\widetilde{\Delta T_j \tau_{21}}|^2\rangle$ and $P_{j,k} = \frac{1}{V}\langle\widetilde{\Delta T_j \tau_{21}}\,\widetilde{\Delta T_k \tau_{21}}^*\rangle$. The first term in Equation (7.31) sums the power spectra of each of the absorbed background sources which is positive and the second term is the sum of their cross power spectra.

We will show that for the range of spatial scales perpendicular to the LoS, accessed by EoR interferometers, the auto power terms in Equation (7.31) dominate the cross power ones at $k_\parallel \gtrsim 10^{-1}$ Mpc^{-1}. We show that the suppression of cross terms is due to two mechanisms: (1) the cross terms are proportional to the cross power spectra

393

between widely separated lines of sight and (2) the cross terms are multiplied by randomly phased sinusoids which cancel out when summed.

7.A.1 The Suppression of the Cross Terms from LoS Cross Power Spectra

To relate the sum in Equation (7.31) to the spectra and locations of the background sources, we assume that all sources are unresolved so that T_j is a delta-function in the plane perpendicular to the LoS. Here, as in McQuinn et al. [139], we will adopt observers coordinates (ℓ, m, ν), rather than comoving coordinates (x, y, z), to emphasize the fact that the the broad-spectrum source does not physically occupy a range of positions along the LoS. In such coordinates, the temperature field of each source can be written as $T_j(\ell, m, \nu)$ where ℓ and m are the direction cosines from the north-south and east-west directions, and ν is the difference from the data cube's central frequency. $\tau_{21}T_j(\ell, m, \nu)$ is given by

$$\tau_{21}T_j(\ell, m, \nu) = \Omega_{pix}\delta(\ell - \ell_j)\delta(m - m_j)\tau_{21}(\ell_j, m_j, \nu)T_j \qquad (7.32)$$

where Ω_{pix} is the solid angle of a map pixel and $\delta(...)$ is the Dirac delta function. For notational simplicity, we will use vector notation to denote direction cosines, $\boldsymbol{\ell} = (\ell, m)$ and their Fourier duals, $\mathbf{u} = (u, v)$. Taking the Fourier transform of $\tau_{21}T_j(\boldsymbol{\ell}, \nu)$ and summing over all sources gets

$$\widetilde{T}_f(\mathbf{u}, \eta) = \sum_j \widetilde{\tau_{21}T_j}(\mathbf{u}, \eta) = \Omega_{pix} \sum_j T_j e^{2\pi i(\boldsymbol{\ell}_j \cdot \mathbf{u})} \int \tau_{21}(\boldsymbol{\ell}_j, \nu)e^{-2\pi i\eta\nu}d\nu. \qquad (7.33)$$

We take the modulus squared of Equation (7.33) and multiply by the cosmology dependent variables, $D_M^2 Y$ [170] that relate observers coordinates to the cosmological comoving coordinates that we've used to define our power spectrum in Equation (7.2).

We find that the sum of the auto terms in Equation (7.31) is

$$\Sigma_{auto} = \frac{D_M^2 \Omega_{pix}^2}{\Omega_{cube}} P_{\tau_{21}}^{LoS}(k_\parallel) \left\langle \sum_j T_j^2 \right\rangle. \tag{7.34}$$

The sum of cross terms is

$$\Sigma_{cross} = 2\frac{D_M^2 \Omega_{pix}^2}{\Omega_{cube}} \sum_{j<k} T_j T_k \Big[\text{Re}\left(P_{\tau_{21};j,k}^{LoS}(k_\parallel) \right) \langle \cos[2\pi(\mathbf{u} \cdot \mathbf{\Delta \ell_{j,k}})] \rangle$$

$$+ \text{Im}\left(P_{\tau_{21};j,k}^{LoS}(k_\parallel) \right) \langle \sin[2\pi(\mathbf{u} \cdot \mathbf{\Delta \ell_{j,k}})] \rangle \Big]$$

$$= 2 P_{\tau_{21}}^{LoS}(k_\parallel)\frac{D_M^2 \Omega_{pix}^2}{\Omega_{cube}} \sum_{j<k} T_j T_k \Bigg[\frac{\text{Re}\left(P_{\tau_{21};j,k}^{LoS}(k_\parallel) \right)}{P_{\tau_{21}}^{LoS}(k_\parallel)} \langle \cos[2\pi(\mathbf{u} \cdot \mathbf{\Delta \ell_{j,k}})] \rangle$$

$$+ \frac{\text{Im}\left(P_{\tau_{21};j,k}^{LoS}(k_\parallel) \right)}{P_{\tau_{21}}^{LoS}(k_\parallel)} \langle \sin[2\pi(\mathbf{u} \cdot \mathbf{\Delta \ell_{j,k}})] \rangle \Bigg], \tag{7.35}$$

where $\mathbf{\Delta \ell_{j,k}} = \mathbf{\ell_j} - \mathbf{\ell_k}$. Here, we define the cross power spectrum between two lines of sight to be

$$P_{\tau_{21};j,k}^{LoS}(k_z) = \frac{1}{L} \int dz dz' e^{ik_z(z-z')} \Delta\tau_{21}(\ell_j, z) \Delta\tau_{21}(\ell_k, z'). \tag{7.36}$$

It is clear from Equation (7.35) that each summand in Σ_{cross} is smaller than each term in Σ_{auto} by a factor of the ratio between the LoS cross power spectra of spatially separated lines of sight and the LoS auto power spectrum. If lines of sight to each source are sufficiently separated, this ratio should be very small. In Figure 7-15 we show the ratios of $\text{Re}\left(P_{\tau_{21};j,k}^{LoS} \right) / P_{\tau_{21}}^{LoS}$ and $\text{Im}\left(P_{\tau_{21};j,k}^{LoS} \right) / P_{\tau_{21}}^{LoS}$ from our fiducial model at $z = 12.2$, separated by $L_\perp = 24$ Mpc which is the mean distance in our data cube between 1000 background sources. Because two sufficiently separated lines of sight should be statistically independent except on the largest spatial scales, these ratios are on the order of $10^{-2} - 10^{-3}$ for $k_\parallel \gtrsim 10^{-1}$ Mpc^{-1} .
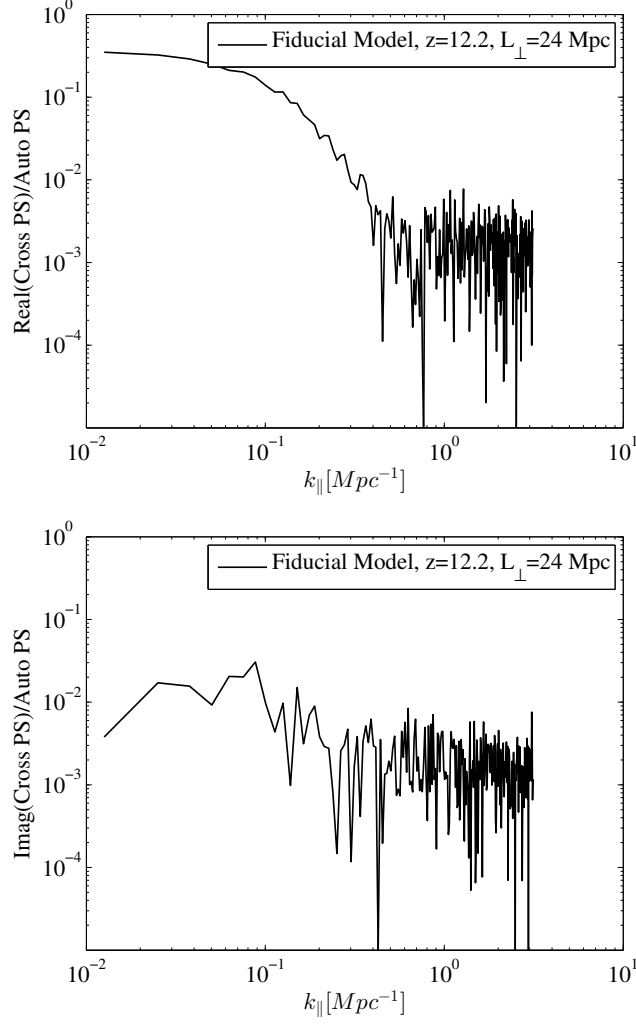
Figure 7-15: The LoS cross power spectra between spatially separated lines of site are on the order of $\sim 100-1000$ times smaller than auto power spectra. In the left figure, we plot the ratio of the real cross power spectra between lines of site separated by 24 Mpc to auto power spectra, and on the right we show the ratio of the imaginary cross power spectrum to the auto power spectrum. In both cases, for $k_\parallel \gtrsim 10^{-1}$ Mpc^{-1}, the cross power spectra are on the order of 10-1000 times smaller. The real cross power spectrum becomes non negligible on scales comparable to the separation between the two lines of site.

7.A.2 Supression of the Cross Terms from Summing the Random Source Phases

The factor of 100-1000 introduced by the ratio of the cross spectra to the auto spectra would be enough to suppress the cross terms if the number of sources were reasonably small. However the number of cross terms relative to the number of auto terms in Equation (7.31) goes as $(N-1)/2$ where N is the number of contributing sources. Thus, even though the cross power spectrum between individual LoS pairs is small, naively summing 100-500 sources could still yield a significant contribution. We now show that summing over many randomly distributed source angles suppresses this.

Since $\mathrm{Im}\left(P^{LoS}_{\tau_{21};j,k}\right)/P^{LoS}_{\tau_{21}}$ is on the same order of, or smaller than $\mathrm{Re}\left(P^{LoS}_{\tau_{21};j,k}\right)/P^{LoS}_{\tau_{21}}$, we will use the real term on both the cosine and sine terms in Equation (7.35) to give an upper bound. Assuming that all sources have the same temperature, $T_j = T_k = T_0$, we may write

$$\Sigma_{cross} \approx 2\sum_{j<k} T_0^2 P^{LoS}_{\tau_{21};j,k}(k_\parallel)\Big[\langle\cos[2\pi(\mathbf{u}\cdot\mathbf{\Delta\ell_{j,k}})]\rangle + \langle\sin[2\pi(\mathbf{u}\cdot\mathbf{\Delta\ell_{j,k}})]\rangle\Big]. \qquad (7.37)$$

Similarly,

$$\Sigma_{auto} \approx N T_0^2 P^{LoS}_{\tau_{21}} \qquad (7.38)$$

Hence the ratio between Σ_{cross} and Σ_{auto} is given by

$$\frac{\Sigma_{cross}}{\Sigma_{auto}} \approx 2\frac{\mathrm{Re}\left(P^{LoS}_{\tau_{21};j,k}(k_\parallel)\right)}{N P^{LoS}_{\tau_{21}}(k_\parallel)}\sum_{j<k}\Big[\langle\cos[2\pi(\mathbf{u}\cdot\mathbf{\Delta\ell_{j,k}})]\rangle + \langle\sin[2\pi(\mathbf{u}\cdot\mathbf{\Delta\ell_{j,k}})]\rangle\Big] \qquad (7.39)$$

Because of the cylindrical symmetry, we need only concern ourselves with a uv cell at $v = 0$ and simply write

$$\Sigma_{cross}/\Sigma_{auto} \approx 2\frac{\mathrm{Re}\left(P^{LoS}_{\tau_{21};j,k}(k_\parallel)\right)}{N P^{LoS}_{\tau_{21}}(k_\parallel)}\sum_{j<k}\Big[\langle\cos[2\pi u_\perp \Delta\ell_{j,k}]\rangle + \langle\sin[2\pi u_\perp \Delta\ell_{j,k}]\rangle\Big]$$

$$= \frac{\mathrm{Re}\left(P^{LoS}_{\tau_{21};j,k}(k_\parallel)\right)}{P^{LoS}_{\tau_{21}}(k_\parallel)}\langle\Sigma_{cos}\left(u_\perp, \Theta, N\right)\rangle, \qquad (7.40)$$

where

$$\Sigma_{cos}(u_\perp, \Theta, N) \equiv \frac{2}{N} \sum_{j<k} \cos[2\pi u_\perp \Delta\ell_{j,k}] + \sin[2\pi u_\perp \Delta\ell_{j,k}]. \qquad (7.41)$$

We can easily compute this ensemble average for any $u_\perp$ by drawing N different source positions distributed randomly over the angular span of the field, Θ, and summing over the sines and cosines of pair-wise angle differences. In Figure 7-16 we show $P[\Sigma_{cos}(u_\perp, \Theta, N)]$ for randomly distributed $\Delta\ell_{j,k}$ for a variety of N, $u_\perp$, and Θ where the minimal $u_\perp$ is set by the maximal scale accessible by an interferometers primary beam, $\sim 1/\Theta$. We calculate these distributions from 10000 random realizations. We see that the distribution of Σ_{cos} is independent of N, Θ, and $u_\perp$ and has a mean of ≈ 0 (which is the quantity that sets the amplitude of Σ_{cross}. As long as sources are randomly distributed, we can expect LoS cross power spectra to suppress the cross terms sum to below the 10% level at $k_\parallel \gtrsim 10^{-1}$ Mpc^{-1}, regardless of the number of terms.

We may finally write.

$$P_f(\mathbf{k}) \approx \sum_j P_j(k_\parallel) = \frac{D_M^2 \Omega_{pix}^2}{\Omega_{cube}} \sum_j T_j^2 P_{\tau_{21}}^{LoS} = \frac{D_M^2 \lambda^4}{4k_B^2 \Omega_{cube}} \sum_j s_j^2 P_{\tau_{21}}^{LoS} \qquad (7.42)$$

where $\lambda = \lambda_{21}(1+z)$ is the wavelength at the center of the data cube, P_j is the absorption power spectrum for the jth source, s_j and T_j are the flux and temperatures of the jth source, Ω_{cube} is the solid angle subtended by the observed volume, and $P_{\tau_{21}}^{LoS}$ is the 1D LoS power spectrum.

We may therefor consider the absorption power spectrum resulting from the forest as simply the sum of the absorption power spectra of each individual source in the background of the source cube. Since all quantities in this sum are positive, we see that the amplitude of the power spectrum increases linearly with the number of sources present behind an observed volume. Because the power spectra for unresolved sources are constant in $k_\perp$, P_f will have a structure that is nearly constant in $k_\perp$.

Hence, for $k \gtrsim 10^{-1}$ Mpc^{-1}, Equation (7.14) simplifies to a sum of the auto power spectra along the LoS to each source. We finish by briefly commenting on the of
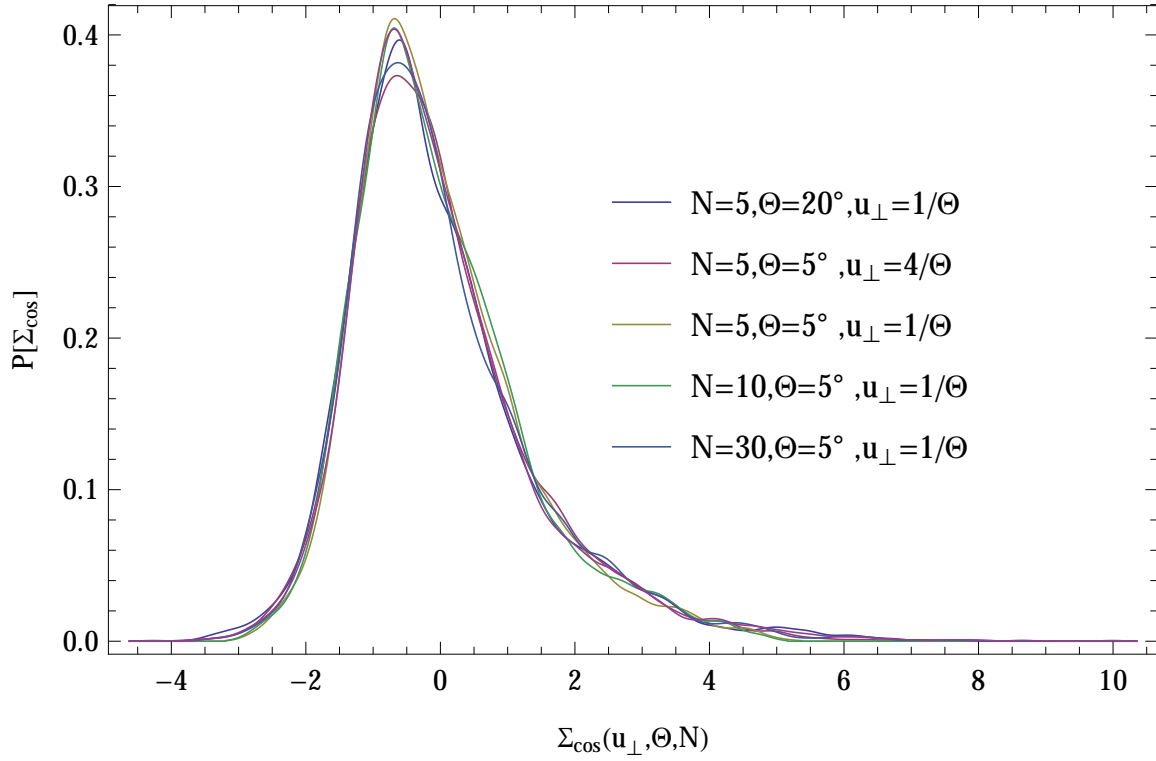
Figure 7-16: Here we see that $P[\Sigma_{cross}(u_\perp, \Theta, N)]$ is invariant in N, Θ, and $u_\perp$, and N with a mean of approximately zero. The lines which indicate, $P[\Sigma_{cross}(u_\perp, \Theta, N)]$, are estimated from 10000 draws. Since $\langle \Sigma_{cross} \rangle \approx 0$ we expect the cross terms to contribute negligibly to P_f in 3D Fourier Space.

the effect of clustering which we have ignored but we find (after comparing Equation (7.15) to our simulations) is still significant. Clustering will cause a disproportionate number of sources to reside in close proximity on the sky. The effect of this is two fold. First, the clustered sources will tend to be behind correlated optical depth columns so that the cross terms between such sources will be better described by auto power spectra. Second, the phases between such sources will be small so that they will not sum to zero. In addition, they will not introduce significant $k_\perp$ structure except at the smallest perpendicular scales. Hence the cross terms introduced by clustered sources will closely resemble the $k_\parallel$ invariant auto terms and simply increase the overall amplitude of P_f. We treat this increase by introducing a multiplicative constant of order unity, A_{cl}, in Equation (7.23).

7.B   Appendix: A Comparison Between Two Source Models

In this paper, we choose to work with the semi-empirical source population in the simulation by Wilman et al. [236]. This choice was in part motivated by the lack of constraints at high redshift and the ease which which we could use data from the W08 simulation using its online interface. Another prediction in the literature for the high redshift radio luminosity function is made by Haiman et al. [85]. This model, like the one in W08, relies on a number of uncertain assumptions but is a more physically motivated bottom up approach which is derived from the cold dark matter power spectrum and assumptions about the black hole-halo mass relation and radio loud fraction. In this appendix, we attempt to understand how our choice of the Wilman source population compares to that in H04. To do this, we attempt to compare the source counts from W08 that contribute the most to P_f to those of H04 who provide cumulative flux counts for $1 - 10$ GHz as a function of redshift. To compare the W08 sources, we compute the percentage of the radio luminosity function integral in Equation (7.16) as a function of the extrapolated $S_{5 \text{ GHz}}$. On the left, in Figure 7-17, a

large fraction of P_f is determined by W08 sources with 5 GHz fluxes between 10 μJy and 10 mJy. We show, in Figure 7-17, the ratio of W08 and H04 source counts with $S_{5\ \mathrm{GHz}}$ between 10 μJy and 10 mJy. The H04 counts fall much faster with redshift than those of W08. At $z \sim 10 - 12$ the number of contributing sources is larger by a factor of ≈ 10 and ≈ 80 by $z \sim 16$.

This comparison is very approximate since different spectral indices are assumed in H04 and W08. However, we emphasize that the observability claims we make in this paper would not apply accurately to the H04 prediction. A more extensive exploration of parameter space will be necessary to determine what range of radio loud source populations may be constrained by the power spectrum technique.

Since P_b is observed to be flat out to $k \approx 10\ \mathrm{Mpc}^{-1}$ while P_f climbs, a more pessimistic source scenario has the effect of pushing the forest dominant region to higher $k_\parallel$ which does not preclude detection with a more powerful telescope such as the SKA.
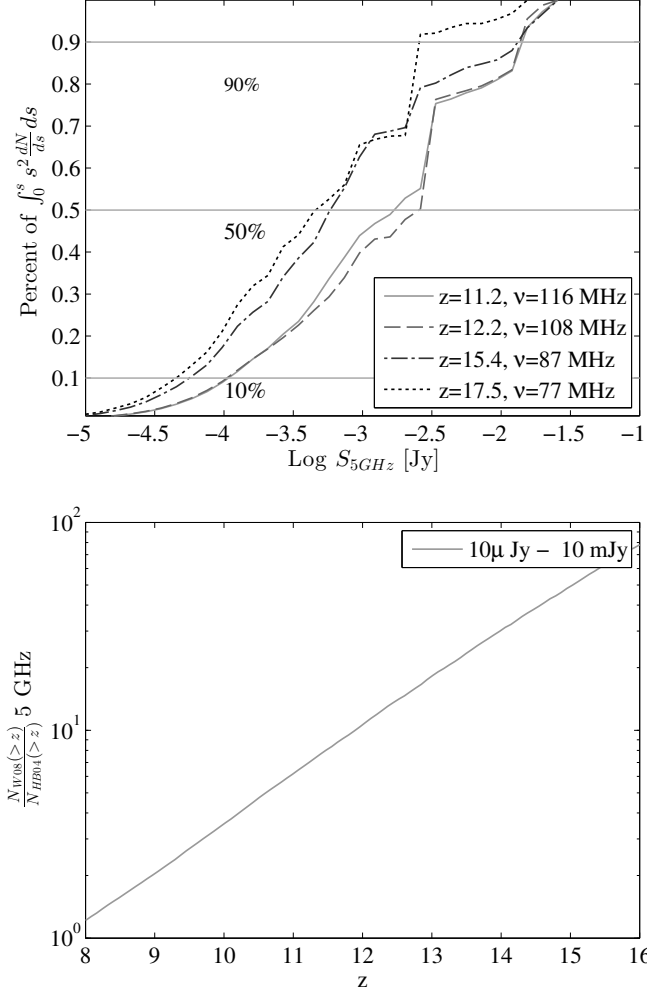
Figure 7-17: Left: The percentage of the integrated luminosity function in Equation (7.16) as a function of the source fluxes at 5GHz for comparison to the catalogue of H04. We see that most contributions to the forest power spectrum come in between $S_{5\ \text{GHz}} = 10\ \mu\text{Jy}$ and $S_{5\ \text{GHz}} = 10\ \text{mJy}$. Right: The ratio of the number of sources with redshift greater than z between $S_{5\ \text{GHz}} = 10\ \mu\text{Jy}$ and $10\ \text{mJy}$ as predicted by the W08 and H04. The W08 simulation over predicts the counts in H04 by a factor ten at $z \gtrsim 12$ and nearly 80 at $z \gtrsim 16$, emphasizing the importance of exploring this widely unconstrained parameter space in future work.

Chapter 8

What Next-Generation 21 cm Power Spectrum Measurements Can Teach Us About the Epoch of Reionization

The content of this chapter was submitted to The Astrophysical Journal *on October 25, 2013 and published [184] as* What Next-Generation 21 cm Power Spectrum Measurements Can Teach Us About the Epoch of Reionization *on January 28, 2014.*

8.1   Introduction

The Epoch of Reionization (EoR) represents a turning point in cosmic history, signaling the moment when large scale structure has become significant enough to impart a global change to the state of the baryonic universe. In particular, the EoR is the period when ultraviolet photons (likely from the first galaxies) reionize the neutral hydrogen in the intergalactic medium (IGM). As such, measurements of the conditions during the EoR promise a wealth of information about the evolution of structure in the universe. Observationally, the redshift of EoR is roughly constrained to be between $z \sim 6$–13, with a likely extended duration; see [74], [188], and [127] for reviews of the field. Given the difficulties of optical/NIR observing at these redshifts, the highly-redshifted 21 cm line of neutral hydrogen has been recognized as a unique

probe of the conditions during the EoR (see Morales and Wyithe [154] and Pritchard and Loeb [188] for recent reviews discussing this technique).

In the last few years, the first generation of experiments targeting a detection of this highly-redshifted 21 cm signal from the EoR has come to fruition. In particular, the LOw Frequency ARray (LOFAR; Yatawatta et al. [244], van Haarlem et al. [229])[1], the Murchison Widefield Array (MWA; Tingay et al. [220], Bowman et al. [29])[2], and the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. [171])[3] have all begun long, dedicated campaigns with the goal of detecting the 21 cm power spectrum. Ultimately, the success or failure of these campaigns will depend on the feasibility of controlling both instrumental systematics and foreground emission. But even if these challenges can be overcome, a positive detection of the power spectrum will likely be marginal at best because of limited collecting area. Progressing from a detection to a characterization of the power spectrum (and eventually, to the imaging of the EoR) will require a next generation of larger 21 cm experiments.

The goal of this paper is to explore the range of constraints that could be achievable with larger 21 cm experiments and, in particular, focus on how those constraints translate into a physical understanding of the EoR. Many groups have analyzed the observable effects of different reionization models on the 21 cm power spectrum; see e.g., Zaldarriaga et al. [246], Furlanetto et al. [72], McQuinn et al. [139], Bowman et al. [25], Bowman et al. [26], Trac and Cen [223], Lidz et al. [117], and Iliev et al. [99]. These studies did not include the more sophisticated understanding of foreground emission that has arisen in the last few years, i.e., the division of 2D cylindrical k-space into the foreground-contaminated "wedge" and the relatively clean "EoR window" [50, 230, 156, 172, 225, 218]. The principal undertaking of this present work is to reconcile these two literatures, exploring the effects of both different EoR histories and foreground removal models on the recovery of astrophysical information from the 21 cm power spectrum. Furthermore, in this work we present some of the first

---

[1]http://www.lofar.org/
[2]http://www.mwatelescope.org/
[3]http://eor.berkeley.edu/

analysis focused on using realistic measurements to distinguish between different theoretical scenarios, rather than simply computing observable (but possibly degenerate) quantities from a given theory. The end result is a set of generic conclusions that both demonstrates the need for a large collecting area next generation experiment and motivates the continued development of foreground removal algorithms.

In order to accomplish these goals, this paper will employ simple models designed to encompass a wide range of possible scenarios. These models are described in Section 8.2, wherein we describe the models for the instrument (Section 8.2.1), foregrounds (Section 8.2.2), and reionization history (Section 8.2.3). In Section 8.3, we present a synthesis of these models and the resultant range of potential power spectrum constraints, including a detailed examination of how well one can recover physical parameters describing the EoR in Section 8.3.5. In Section 8.4, we conclude with several generic messages about the kind of science the community can expect from 21 cm experiments in the next ~ 5 years.

8.2   The Models

In this section we present the various models for the instrument (Section 8.2.1), foreground removal (Section 8.2.2), and reionization history (Section 8.2.3) used to explore the range of potential EoR measurements. In general, these models are chosen not because they necessarily mirror specific measurements or scenarios, but rather because of their simplicity while still encompassing a wide range of uncertainty about many parameters. We choose several different parameterizations of the foreground removal algorithms, and use simple simulations to probe a wide variety of reionization histories. Our model telescope (described below in Section 8.2.1) is based off the proposed Hydrogen Epoch of Reionization Array (HERA); we present sensitivity calculations and astrophysical constraints for other 21 cm experiments in the appendix.
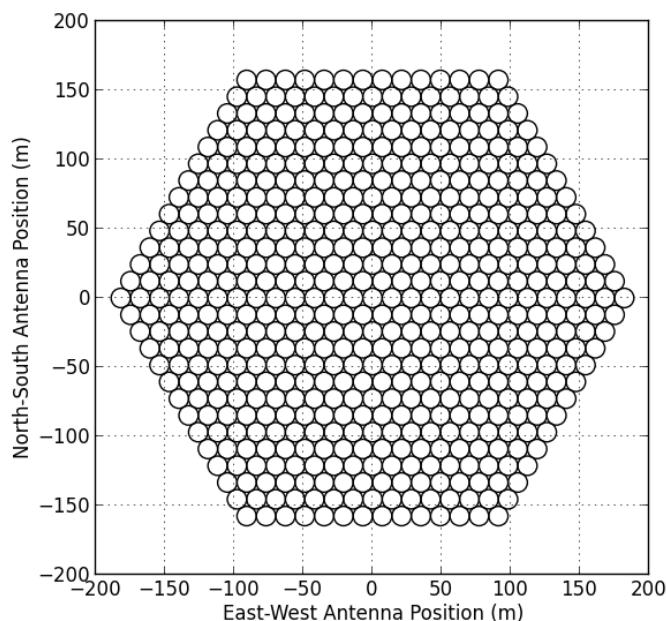
Figure 8-1: The 547-element, hexagonally packed HERA concept design, with 14 m reflector elements. Outrigger antennas may be included in the final design for the purposes of foreground imaging, but they are not treated here, since they add little to power spectrum sensitivity.

8.2.1   The Telescope Model

The most significant difference between the current and next generations of 21 cm instruments will be a substantial increase in collecting area and, therefore, sensitivity. In the main body of this work, we use an instrument modeled after a concept design for the Hydrogen Epoch of Reionization Array (HERA)[4]. This array consists of 547 zenith-pointing 14 m diameter reflecting-parabolic elements in a close-packed hexagon, as shown in Figure 8-1. The total collecting area of this array is $84,000$ m^2, or approximately one tenth of a square kilometer. The goal of this work is not to justify this particular design choice, but rather to show that this scale instrument enables a next level of EoR science beyond the first generation experiments. In the appendix, we present the resultant sensitivities and achievable constraints on the astrophysical parameters of interest for several other 21 cm telescopes: PAPER, the MWA, LOFAR, and a concept design for the SKA-Low Phase 1. Generically, we

---

[4]http://reionization.org/

Observing Frequency	50–225 MHz
T_{receiver}	100 K
Parabolic Element Size	14 m
Number of Elements	547
Primary beam size	8.7° FWHM at 150 MHz
Configuration	Close-packed hexagonal
Observing mode	Drift-scanning at zenith

Table 8.1: Fiducial System Parameters.

find that power spectrum sensitivities are a strong function of array configuration, especially compactness and redundancy. However, once the power spectrum sensitivity of an array is known, constraints on reionization physics appear to be roughly independent of other paramters.

In many ways, the HERA concept array design is quite representative of 21 cm EoR experiments over the next $\sim$ 5–10 years. As mentioned, it has a collecting area of order a tenth of a square kilometer — significantly larger than any current instrument, but smaller than Phase 1 of the low-frequency Square Kilometre Array (SKA1-low)[5]. (See Table 8.4 for a summary of different EoR telescopes.) In terms of power spectrum sensitivity, [170] demonstrated the power of array redundancy for reducing thermal noise uncertainty, and showed that a hexagonal configuration has the greatest instantaneous redundancy. In this sense, the HERA concept design is optimized for power spectrum measurements. Other configurations in the literature have been optimized for foreground imaging or other additional science; the purpose of this work is not to argue for or against these designs. Rather, we concentrate primarily on science with the 21 cm power spectrum, and use the HERA concept design as representative of power spectrum-focused experiments. Obviously, arrays with more (less) collecting area will have correspondingly greater (poorer) sensitivity. The key parameters of our fiducial concept array are given in Table 8.1, and constraints achievable with other arrays are presented in the appendix.

---

[5]http://www.skatelescope.org/

8.2.1.1 Calculating Power Spectrum Sensitivity

To calculate the power spectrum sensitivity of our fiducial array, we use the method presented in [183], which is briefly summarized here. This method begins by creating the uv coverage of the observation by gridding each baseline into the uv plane, including the effects of earth-rotation synthesis over the course of the observation. We choose uv pixels the size of the antenna element in wavelengths, and assume that any baseline samples only one pixel at a time. Each pixel is treated as an independent sample of one $k_\perp$-mode, along which the instrument samples a wide range of $k_\parallel$-modes specified by the observing bandwidth. The sensitivity to any one mode of the dimensionless power spectrum is given by Equation (4) in [183], which is in turn derived from Equation (16) of [170]:

$$\Delta_{\mathrm{N}}^2(k) \approx X^2 Y \frac{k^3}{2\pi^2} \frac{\Omega'}{2t} T_{\mathrm{sys}}^2, \tag{8.1}$$

where $X^2 Y$ is a cosmological scalar converting observed bandwidths and solid angles to $h\mathrm{Mpc}^{-1}$, $\Omega' \equiv \Omega_{\mathrm{p}}^2/\Omega_{\mathrm{pp}}$ is the solid angle of the power primary beam (Ω_{p}) squared, divided by the solid angle of the square of the power primary beam (Ω_{pp}),[6] t is the integration time on that particular k-mode, and T_{sys} is the system temperature. It should also be noted that this equation is dual-polarization, i.e., it assumes both linear polarizations are measured simultaneously and then combined to make a power spectrum estimate. Similar forms of this equation appear in [151] and [139] which differ only by the polarization factor and power-squared primary beam correction.

In our formalism, each measured mode is attributed a noise value calculated from Equation 8.1 (see Section 8.2.1.2 for specifics on the values of each parameter). Independent modes can be combined in quadrature to form spherical or cylindrical power spectra as desired. One has a choice of how to combine non-instantaneously redundant baselines which do in fact sample the same $k_\perp/uv$ pixel. Such a situation can arise either through the effect of the gridding kernel creating overlapping uv footprints

---

[6]Although [170] and [183] originally derived this relation with the standard power primary beam Ω, it was shown in [173] that the power-squared beam enters into the correct normalizing factor.

on similar length baselines ("partial coherence"; Hazelton et al. [89]), or through the effect of earth-rotation bringing a baseline into a uv pixel previously sampled by another baseline. Naïvely, this formalism treats these samples as perfectly coherent, i.e., we add the integration time of each baseline within a uv pixel. As suggested by [89], however, it is possible that this kind of simple treatment could lead to foreground contamination in a large number of Fourier modes. To explore the ramifications of this effect, we will also consider a case where only baselines which are instantaneously redundant are added coherently, and all other measurements are added in quadrature when binning. We discuss this model more in Section 8.2.2.

Since this method of calculating power spectrum sensitivities naturally tracks the number of independent modes measured, sample variance is easily included when combining modes by adding the value of the cosmological power spectrum to each (u, v, η)-voxel (where η is the line-of-sight Fourier mode) before doing any binning. (Note that in the case where only instantaneously redundant baselines are added coherently, partially coherent baselines do not count as independent samples for the purpose of calculating sample variance.) Unlike [183], we do not include the effects of redshift-space distortions in boosting the line of sight signal, since they will not boost the power spectrum of ionization fluctuations, which is likely to dominate the 21 cm power spectrum at these redshifts. We also ignore other second order effects on the power spectrum, such as the "light-cone" effect [51, 113].

8.2.1.2 Telescope and Observational Parameters

For the instrument value of T_{sys} we sum a frequency independent 100 K receiver temperature with a frequency dependent sky temperature, $T_{\mathrm{sky}} = 60\mathrm{K} \ (\lambda/1 \ \mathrm{m})^{2.55}$ [216], giving a sky temperature of 351 K at 150 MHz. Although this model is $\sim$ 100 K lower than the system measured by [173], it is consistent with recent LOFAR measurements [244, 229]. Since the smaller field of view of HERA will lead to better isolation of a Galactic cold patch, we choose this empirical relation for our model.

For the primary beam, we use a simple Gaussian model with a Full-Width Half-Max (FWHM) of $1.06\lambda/D = 8.7°$ at 150 MHz. We assume the beam linearly evolves

in shape as a function of frequency. In the actual HERA instrument design, the PA-PER dipole serves as a feed to the larger parabolic element. Computational E&M modeling suggests this setup will have a beam with FWHM of 9.8°. Furthermore, the PAPER dipole response is specifically designed to evolve more slowly with frequency than our linear model. Although the frequency dependence of the primary beam enters into our sensitivity calculations in several places (including the pixel size in the uv plane), the dominant effect is to change the normalization of the noise level in Equation 8.1. For an extreme case with no frequency evolution in the primary beam size (relative to 150 MHz), we find that the resultant sensitivities increase by up to 40% at 100 MHz (due to a smaller primary beam than the linear evolution model), and decrease by up to 30% at 200 MHz (due to larger beam). While all instruments will have some degree of primary beam evolution as a function of frequency, this extreme model demonstrates that some of the poor low-frequency (high-redshift) sensitivities reported below can be partially mitigated by a more frequency-independent instrument design (although at the expense of sensitivity at higher frequencies).

It should be pointed out that for snap-shot observations, the large-sized HERA dishes prevent measurements of the largest transverse scales. At 150 MHz ($z = 8.5$), the minimum baseline length of 14 m corresponds to a transverse k-mode of $k_\perp = 0.0068h\mathrm{Mpc}^{-1}$. This array will be unable to observe transverse modes on larger scales, without mosaicing or otherwise integrating over longer than one drift through the primary beam. The sensitivity calculation used in this work does not account for such an analysis, and therefore will limit the sensitivity of the array to larger-scale modes. For an experiment targeting unique cosmological information on the largest cosmic scales (e.g. primordial non-Gaussianity), this effect may prove problematic. For studies of the EoR power spectrum, the limitation on measurements at low $k_\perp$ has little effect on the end result, especially given the near ubiquitous presence of foreground contamination on large-scales in our models (Section 8.2.2).

The integration time t on a given k mode, is determined by the length of time any baseline in the array samples each uv pixel over the course of the observation. Since we assume a drift-scanning telescope, the length of the observation is set by the size

of the primary beam. The time it takes a patch of sky to drift through the beam is the duration over which we can average coherently. For the $\sim 10°$ primary beam model above, this time is ~ 40 minutes.

We assume that there exists one Galactic "cold patch" spanning 6 hours in right ascension suitable for EoR observations, an assumption which is based on measurements from both PAPER and the MWA and on previous models (e.g. de Oliveira-Costa et al. [54]). There are thus 9 independent fields of 40 minutes in right ascension (corresponding to the primary beam size calculated above) which are observed per day. We also assume EoR-quality observations can only be conducted at night, yielding ~ 180 days per year of good observing. Therefore, our thermal noise uncertainty (i.e. the 1σ error bar on the power spectrum) is reduced by a factor of $\sqrt{9 \times 180}$ over that calculated from one field, whereas the contribution to the errors from sample variance is only reduced by $\sqrt{9}$.

8.2.2 Foregrounds

Because of its spectral smoothness, foreground emission is expected to contaminate low order line-of-sight Fourier modes in the power spectrum. Of great concern, though, are chromatic effects in an interferometer's response, which can introduce spectral structure into foreground emission. However, recent work has shown that these chromatic mode-mixing effects do not indiscriminately corrupt all the modes of the power spectrum. Rather, foregrounds are confined to a "wedge"-shaped region in the 2D $(k_\perp, k_\parallel)$ plane, with more $k_\parallel$ modes free from foreground contamination on the shortest baselines (i.e. at the smallest $k_\perp$ values) [50, 230, 156, 172, 225], as schematically diagrammed in Figure 8-2. Power spectrum analysis in both [59] and [182] reveal the presence of the wedge in actual observations. The single-baseline approach [172] used in [182] yields a cleaner EoR window, although at the loss of some sensitivity that comes from combining non-redundant baselines.

However, there is still considerable debate about where to define the "edge" of the wedge. Our three foreground models — summarized in Table 8.2 — differ in their choice of "wedge edge." Our pessimistic model also explores the possibility
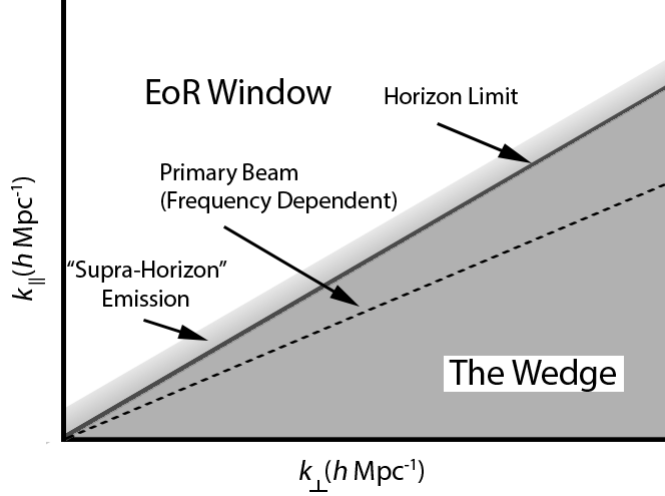
Figure 8-2: A schematic diagram of the wedge and EoR window in 2D k-space. See Section 8.2.2 for explanations of the terms.

that systematic effects discussed in [89] could prevent to coherent addition of partially redundant baselines. It should be noted that although we use the shorthand "foreground model" to describe these three scenarios, in many ways these represent *foreground removal models*, since they pertain to improvements over current analysis techniques that may better separate foreground emission from the 21 cm signal.

8.2.2.1 Foreground Removal Models

At present, observational limits on the "edge" to the foreground wedge in cylindrical $(k_\perp, k_\parallel)$-space are still somewhat unclear. [182] find the wedge to extend as much as $\Delta k_\parallel = 0.05$–$0.1\ h\mathrm{Mpc}^{-1}$ beyond the "horizon limit," i.e., the $k_\parallel$ mode on a given baseline that corresponds to the chromatic sine wave created by a flat-spectrum source of emission located at the horizon. (This mode in many ways represents a fundamental limit, as the interference pattern cannot oscillate any faster for a flat-spectrum source of celestial emission; see Parsons et al. [172] for a full discussion of the wedge in the language of geometric delay space.) Mathematically, the horizon limit is:

$$k_{\parallel,\mathrm{hor}} = \frac{2\pi}{Y}\frac{|\vec{b}|}{c} = \left(\frac{1}{\nu}\frac{X}{Y}\right) k_\perp, \qquad (8.2)$$

where $|\vec{b}|$ is the baseline length in meters, c is the speed of light, ν is observing frequency, and X and Y are the previously described cosmological scalars for converting observed bandwidths and solid angles to $h\mathrm{Mpc}^{-1}$, respectively, defined in [170] and [74]. [182] attribute the presence of "supra-horizon" emission — emission at $k_\parallel$ values greater than the horizon limit — to spectral structure in the foregrounds themselves, which creates a convolving kernel in k-space. [172] predict that the wedge could extend as much as $\Delta k_\parallel = 0.15\ h\mathrm{Mpc}^{-1}$ beyond the horizon limit at the level of the 21 cm EoR signal. This supra-horizon emission has a dramatic effect on the size of the EoR window, increasing the $k_\parallel$ extent of the wedge by nearly a factor of 4 on the 16λ-baselines used by PAPER in [173].

Others have argued that the wedge will extend not to the horizon limit, but only to the edges of the field-of-view, outside of which emission is too attenuated to corrupt the 21 cm signal. If achievable, this smaller wedge has a dramatic effect on sensitivity, since theoretical considerations suggest that signal-to-noise decreases quickly with increasing $k_\perp$ and $k_\parallel$. If one compares the sensitivity predictions in [172] for PAPER-132 and [15] for MWA-128 (two comparably sized arrays), one finds that these two different wedge definitions account for a large portion of the difference between a marginal 2σ EoR detection and a 14σ one.

While clearly inconsistent with the current results in [182], such a small wedge may be achievable with new advances in foreground subtraction. A large literature of work has gone into studying the removal of foreground emission from 21 cm data (e.g. Morales et al. [155], Bowman et al. [28], Liu et al. [122], Liu and Tegmark [120], Chapman et al. [39], Dillon et al. [58], Chapman et al. [40]). If successful, these techniques offer the promise of working within the wedge. However, despite the huge sensitivity boost, working within the wedge clearly presents additional challenges beyond simply working within the EoR window. Working within the EoR window requires only keeping foreground leakage from within the wedge to a level below the 21 cm signal; the calibration challenge for this task can be significantly reduced by techniques which are allowed to remove EoR signal from within the wedge [173]. Working within the wedge requires foreground removal with up to 1 part in 10^{10}

413

accuracy (in mK^2) while leaving the 21 cm signal unaffected. Ensuring that calibration errors do not introduce covariance between modes is potentially an even more difficult task. Therefore, given the additional effort it will take to be convinced that a residual excess of power post-foreground subtraction can be attributed to the EoR, it seems plausible that the first robust detection and measurements of the 21 cm EoR signal will come from modes outside the wedge.

To further complicate the issue, several effects have been identified which can scatter power from the wedge into the EoR window. [150] demonstrate how combining redundant visibilities without image plane correction (as done by PAPER) can corrupt the EoR signal outside the wedge, due to the effects of instrumental polarization leakage. [150] predict a level of contamination based on simulations of the polarized point source population at low frequencies. Although this predicted level of contamination may already be ruled out by measurements from [20], these effects are a real concern for 21 cm EoR experiments. In the present analysis, however, we do not consider this contamination; rather, we assume that the dense uv coverage of our concept array will allow for precision calibration and image-based primary beam correction not possible with the sparse PAPER array. Through careful and concerted effort this systematic should be able to be reduced to below the EoR level.

As discussed in Section 8.2.1.1, we do consider the "multi-baseline mode mixing" effects presented in [89]. These effects may result when partially coherent baselines are combined to improve power spectrum sensitivity, introducing additional spectral structure in the foregrounds and thus complicating their mitigation. Conversely, the fact that only instantaneously redundant baselines were combined in [182] and [173] was partially responsible for the clear separation between the wedge and EoR window. Since recent, competitive upper limits were set using this conservative approach, we include it as our "pessimistic" foreground strategy, noting that recent progress in accounting for the subtleties in partially coherent analyses [89] make it likely that better schemes will be available soon.

To encompass all these uncertainties in the foreground emission and foreground removal techniques, we use three models for our foregrounds, which we refer to in

Model	Parameters
Moderate	Foreground wedge extends 0.1 $h\mathrm{Mpc}^{-1}$ beyond horizon limit
Pessimistic	Foreground wedge extends 0.1 $h\mathrm{Mpc}^{-1}$ beyond horizon limit, and only instantaneously redundant baselines can be combined coherently
Optimistic	Foreground wedge extends to FWHM of primary beam

Table 8.2: Summary of the three foreground removal models.

shorthand as "pessimistic," "moderate," and "optimistic". These models are summarized in Table 8.2.

The "moderate" model is chosen to closely mirror the predictions and data from PAPER. In this model the wedge is considered to extend $\Delta k_\parallel = 0.1 \ h\mathrm{Mpc}^{-1}$ beyond the horizon limit. The exact scale of the "horizon+.1" limit to the wedge is motivated by the predictions of [172] and the measurements of [182] and [173]. Although the exact extent of the "supra-horizon" emission (i.e. the "+.1") at the level of the EoR signal remains to be determined, all of these constraints point to a range of 0.05 to 0.15 $h\mathrm{Mpc}^{-1}$. The uncertainty in this value does not have a large effect on the ultimate power spectrum sensitivity of next generation measurements. For shorthand, we will sometimes refer this model as having a "horizon wedge."

The "pessimistic" model uses the same horizon wedge as the moderate model, but assumes that only instantaneously redundant baselines are coherently combined. Any non-redundant baselines which sample the same uv pixel as another baseline — either through being similar in length and orientation or through the effects of earth rotation — are added incoherently. In effect, this model covers the case where the multi-baseline mode-mixing of [89] cannot be corrected for. Significant efforts are underway to develop pipelines which correct for this effect and recover the sensitivity boost of partial coherence; since these algorithms have yet to be demonstrated on actual observations, however, we consider this our pessimistic scenario.

The final "optimistic" model, assumes the EoR window remains workable down to $k_\parallel$ modes bounded by the FWHM of the primary beam, as opposed to the horizon: $k_{\parallel,\mathrm{pb}} = \sin(\mathrm{FWHM}/2) \times k_{\parallel,\mathrm{hor}}$. The specific choice of the FWHM is somewhat arbitrary; one could also consider a wedge extending the first-null in the primary beam

(although this is ill-defined for a Gaussian beam model). Alternatively, one might envision a "no wedge" model meant to mirror the case where foreground removal techniques work optimally, removing all foreground contamination down to the intrinsic spectral structure of the foreground emission. In practice, the small $\sim 10°$ size of the HERA primary beam renders these different choices effectively indistinguishable. Therefore, our choice of the primary beam FWHM can also be considered representative of nearly all cases where foreground removal proves highly effective. As of the writing of this paper, no foreground removal algorithms have proven successful to these levels, although this is admittedly a somewhat tautological statement, since no published measurements have reached the sensitivity level of an EoR detection. Furthermore, the sampling point-spread function (PSF) in k-space at low k's is expected to make clean, unambiguous retrieval of these modes exceedingly difficult [120, 172], although the small size of the HERA primary beam ameliorates this problem by limiting the scale of this PSF. We find this effect to represent a small ($\lesssim 5\%$) correction to the low-k sensitivities reported in this work. In effect, the optimistic model is included to both show the effects of foregrounds on the recovery of the 21 cm power spectrum, and to give an impression of what could be achievable. For shorthand, this model will be referred to as having a "primary beam wedge."

Incorporating these foreground models into the sensitivity calculations described in Section 8.2.1 is quite straightforward. Modes deemed "corrupted" by foregrounds according to a model are simply excluded from the 3D k-space cube, and therefore contribute no sensitivity to the resultant power spectrum measurements.

8.2.3   Reionization

In order to encompass the large theoretical uncertainties in the cosmic reionization history, we use the publicly available `21cmFAST`[7] code v1.01 [143, 144]. This semi-numerical code allows us to quickly generate large-scale simulations of the ionization field (400 Mpc on a side) while varying key parameters to examine the possible variations in the 21 cm signal. Following [145], we choose three key parameters to

---

[7]http://homepage.sns.it/mesinger/DexM___21cmFAST.html/

encompass the maximum variation in the signal:

1. ζ, *the ionizing efficiency*: ζ is a conglomeration of a number of parameters relating to the amount of ionizing photons escaping from high-redshift galaxies: f_{esc}, the fraction of ionizing photons which escape into the IGM, f_*, the star formation efficiency, N_γ, the number of ionizing photons produced per baryon in stars, and n_{rec} the average number of recombinations per baryon. Rather than parameterize the uncertainty in these quantities individually, it is common to define $\zeta = f_{\text{esc}} f_* N_\gamma / (1 + n_{\text{rec}})$ [72]. We explore a range of $\zeta = 10 - 50$ in this work, which is generally consistent with current CMB and Lyα constraints on reionization [145].

2. T_{vir}, *the minimum virial temperature of halos producing ionizing photons*: T_{vir} parameterizes the mass of the halos responsible for reionization. Typically, T_{vir} is chosen to be 10^4 K, which corresponds to a halo mass of 10^8 M$_\odot$ at $z = 10$. This value is chosen because it represents the temperature at which atomic cooling becomes efficient. In this work, we explore T_{vir} ranging from 10^3–3×10^5 K to span the uncertainty in high-redshift galaxy formation physics as to which halos host significant stellar populations (see e.g. Haiman et al. [84], Abel et al. [1] and Bromm et al. [31] for lower mass limits on star-forming halos, and e.g. Mesinger and Dijkstra [142] and Okamoto et al. [165] for feedback effects which can suppress low mass halo star formation).

3. R_{mfp}, *the mean free path of ionizing photons through the intergalactic medium (IGM)*: R_{mfp} sets the maximum size of HII regions that can form during reionization. Physically, it is set by the space density of Lyman limit systems, which act as sinks of ionizing photons. In this work, we explore a range of mean free paths from 3 to 80 Mpc, spanning the uncertainties in current measurements of the mean free path at $z \sim 6$ [202].

We note there are many other tunable parameters that could affect the reionization history. In particular, the largest 21 cm signals can be produced in models where the

IGM is quite cold during reionization (cf. Parsons et al. [173]). We do not include such a model here, and rather focus on the potential uncertainties within "vanilla" reionization; for an analysis of the detectability of early epochs of X-ray heating, see [42] and [147]. Also note that `21cmFAST` assumes the values of the EoR parameters are constant over all redshifts considered. With the exception of our three EoR variables, we use the fiducial parameters of the `21cmFAST` code; see [144] for more details.

Note we do assume that $T_{\text{spin}} \gg T_{\text{CMB}}$ at all epochs, which could potentially create a brighter signal at the highest redshifts. Given that thermal noise generally dominates the signal at the highest redshifts regardless, we choose to ignore this effect, noting that it will only increase the difficulties of $z > 10$ observations we describe below. (Although this situation may be changed by the alternate X-ray heating scenarios considered in Mesinger et al. [147].)

8.2.3.1 "Vanilla" Model

For the sake of comparison, it is worthwhile to have one fiducial model with "middle-ground" values for all the parameters in question. We refer to this model as our "vanilla" model. Note that this model was not chosen because we believe it most faithfully represents the true reionization history of the universe (though it is consistent with current observations). Rather, it is simply a useful point of comparison for all the other realizations of the reionization history. In this model, the values of the three parameters being studied are $\zeta = 31.5$, $T_{\text{vir}} = 1.5 \times 10^4$ K and $R_{\text{mfp}} = 30$ Mpc. This model achieves 50% ionization at $z \sim 9.5$, and complete ionization at $z \sim 7$. The redshift evolution of the power spectrum in this model is shown in Figure 8-3.

8.2.3.2 The Effect of the Varying the EoR Parameters

The effects of varying ζ, T_{vir} and R_{mfp} are illustrated in Figure 8-4. Each row shows the effect of varying one of the three parameters while holding the other two fixed. The middle panel in each row is for our vanilla model, and thus is the same as Figure 8-3 (although the $z = 8$ curve is not included for clarity). Several qualitative observations can immediately be made. Firstly, we can see from the top row that ζ
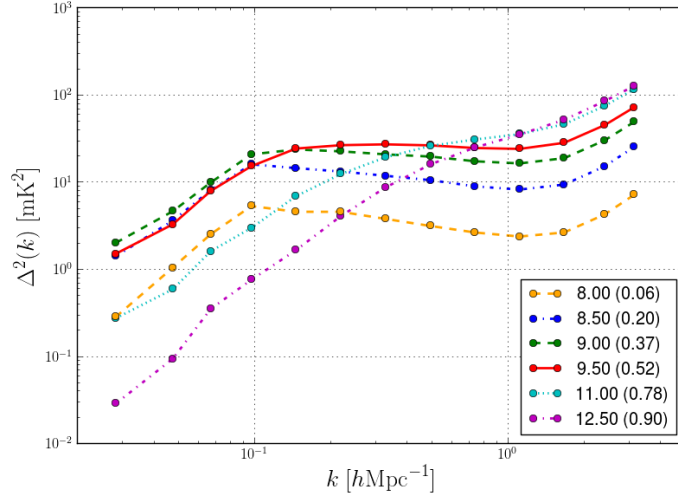
Figure 8-3: Power spectra at several redshifts for our vanilla reionization model with $\zeta = 31.5$, $T_{\mathrm{vir}} = 1.5 \times 10^4$ K, and $R_{\mathrm{mfp}} = 30$ Mpc. Numbers in parentheses give the neutral fraction at that redshift.

does not significantly change the shape of the power spectrum, but only the duration and timing of reionization. This is expected, since the same sources are responsible for driving reionization regardless of ζ. Rather, it is only the number of ionizing photons that these sources produce that varies.

Secondly, we can see from the middle row that the most dramatic effect of T_{vir} is to substantially change the timing of reionization. Our high and low values of T_{vir} create reionization histories that are inconsistent with current constraints from the CMB and Lyα forest [65, 92]. This alone does not rule out these values of T_{vir} for the minimum mass of reionization galaxies, but it does mean that some additional parameter would have to be adjusted within our vanilla model to create reasonable reionization histories. We can also see that the halo virial temperature affects the shape of the power spectrum. When the most massive halos are responsible for reionization, we see significantly more power on very large scales than in the case where low-mass galaxies reionize the universe.

Finally, the bottom row shows that the mean free path of ionizing photons also affects the amount of large scale power in the 21 cm power spectrum. R_{mfp} values of 30 and 80 Mpc produce essentially indistinguishable power spectra, except at the very largest scales at late times. However, the very small value of R_{mfp} completely
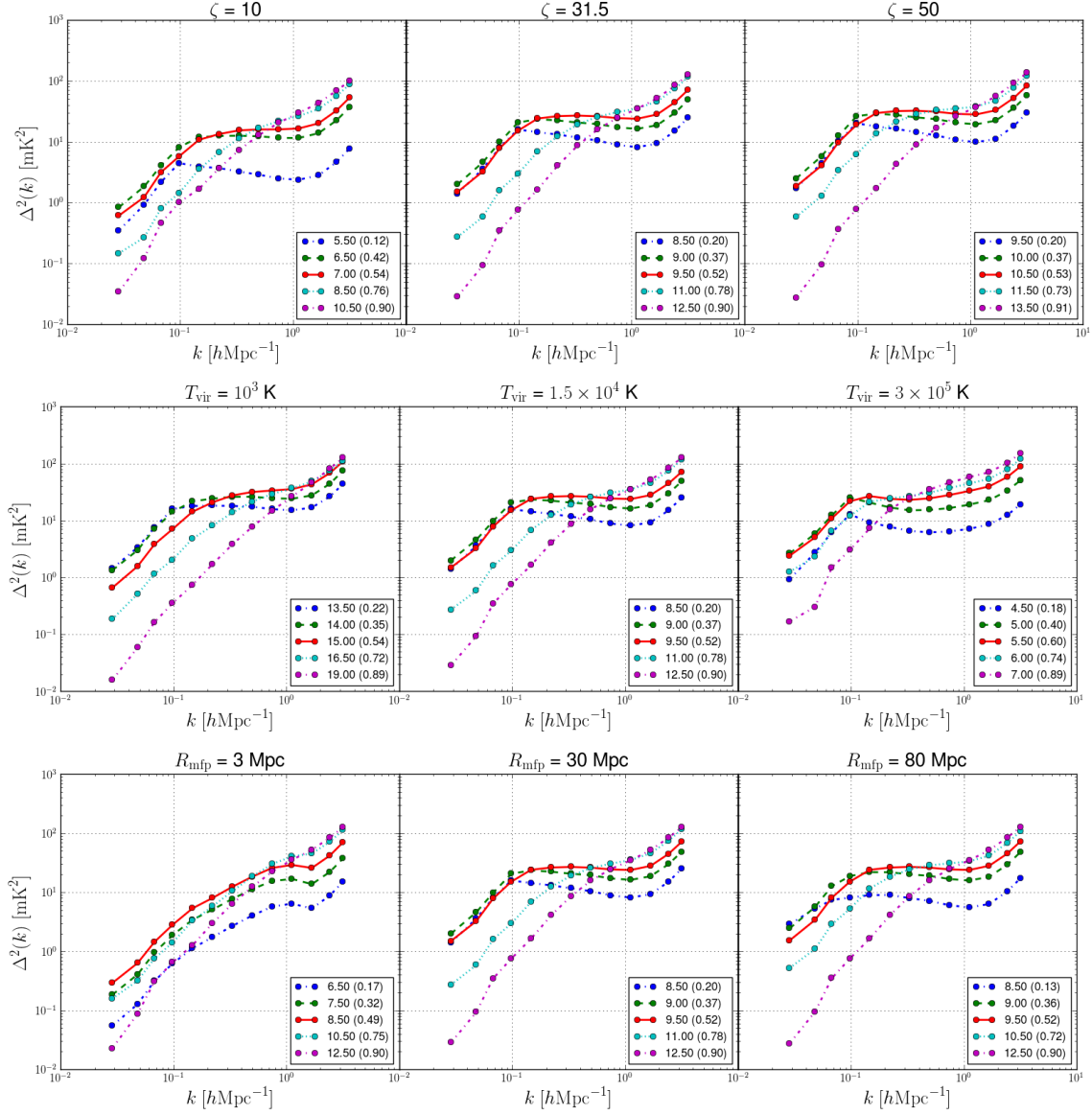
Figure 8-4: Power spectra as a function of redshift for the low, high, and fiducial values of the ionizing efficiency, ζ (top row), T_{vir} (middle row) and R_{mfp} (bottom row). Exact values of each parameter are given in the panel title. Numbers in parentheses give the neutral fraction at that redshift. The central panel is the vanilla model and is identical to Figure 8-3 (although the $z = 8$ curve is not included for clarity). Colors in each panel map to roughly the same neutral fraction. Qualitative effects of varying each parameter are apparent: ζ changes the timing of reionization but not the shape of the power spectrum; T_{vir} drastically alters the timing of reionization with smaller effects on the power spectrum shape; and small values of R_{mfp} reduce the amount of low k power.

changes the shape of the power spectrum, resulting in a steep slope versus k, even at 50% ionization, where most models show a fairly flat power spectrum up to a "knee" feature on larger scales. In section 8.3.4, we consider using some of these characteristic features to qualitatively assess properties from reionization in 21 cm measurements.

8.3 Results

In this section, we will present the predicted sensitivities that result from combinations of EoR and foreground models. We will focus predominantly on the moderate model where one can take advantage of partially-redundant baselines, but the wedge still contaminates $k_\parallel$ modes below 0.1 $h\mathrm{Mpc}^{-1}$ above the horizon limit.' In presenting the sensitivity levels achievable under the other two foreground models, we focus on the additional science that will be prevented/allowed if these models represent the state of foreground removal analysis.

We will take several fairly orthogonal approaches towards understanding the science that will be achievable. First, in Section 8.3.1, our approach is to attempt to cover the broadest range of possible power spectrum shapes and amplitudes in order to make generic conclusions about the detectability of the 21 cm power spectrum. In Section 8.3.2, Section 8.3.3, and Section 8.3.4, we focus on our vanilla reionization model and semi-quantitatively explore the physical lessons the predicted sensitivities will permit. Finally, in Section 8.3.5, we undertake a Fisher matrix analysis and focus on specific determinations of EoR parameters with respect to the fiducial evolution of our vanilla model, exploring the degeneracies between parameters and providing lessons in how to break them. The end result of these various analyses is a holistic picture about the kinds of information we can derive from next generation EoR measurements.

8.3.1 Sensitivity Limits

In this section, we consider the signal-to-noise ratio of power spectrum measurements achievable under our various foreground removal models. The main results are pre-

sented in Figures 8.3.1.2, 8-6, 8-7, and 8-8. Figure 8.3.1.2 shows the constraints on the 50% ionization power spectrum in our vanilla model for each of the three foreground models, as well as the measurement significances of alternate ionization histories using the vanilla model. Figures 8-6, 8-7, and 8-8 show the power spectrum measurement signifances that result when the EoR parameters are varied for the moderate, pessimistic, and optimistic foreground models respectively.

8.3.1.1 Methodology

In order to explore the largest range of possible power spectrum shapes and amplitudes, it is important to keep in mind the small but non-negligible spread between various theoretical predictions in the literature. To avoid having to run excessive numbers of simulations, we make use of the observation that much of the differences between simulations is due to discrepancies in their predictions for the ionization history $x_{\mathrm{HI}}(z)$, in the sense that the differences decrease if neutral fraction (rather than redshift) is used as the time coordinate for cross-simulation comparisons. We thus make the ansatz that given a single set of parameters $(\zeta, T_{\mathrm{vir}}, R_{\mathrm{mfp}})$, the 21cmFAST power spectrum outputs can (modulo an overall normalization factor) be translated in redshift to mimic the outputs of alternative models that predict a different ionization history. In practice, the 21cmFAST simulation provides a suite of power spectra in either (a) fixed steps in z or (b) approximately fixed steps in x_{HI}, but constrained to appear at a single z. We utilize the latter set, and "extrapolate" each neutral fraction to a variety of redshifts by scaling the amplitude with the square of the mean temperature of the IGM as $(1+z)$, as anticipated when ionization fractions dominate the power spectrum [139, 117]. While not completely motivated by the physics of the problem (since within 21cmFAST a given set of EoR parameters does produce only one reionization history), this approach allows us to explore an extremely wide range of power spectrum amplitudes while running a reasonable number of simulations.

8.3.1.2 Moderate Foreground Model

Figure 8.3.1.2 shows forecasts for constraints on our fiducial reionization model under the three foreground scenarios.
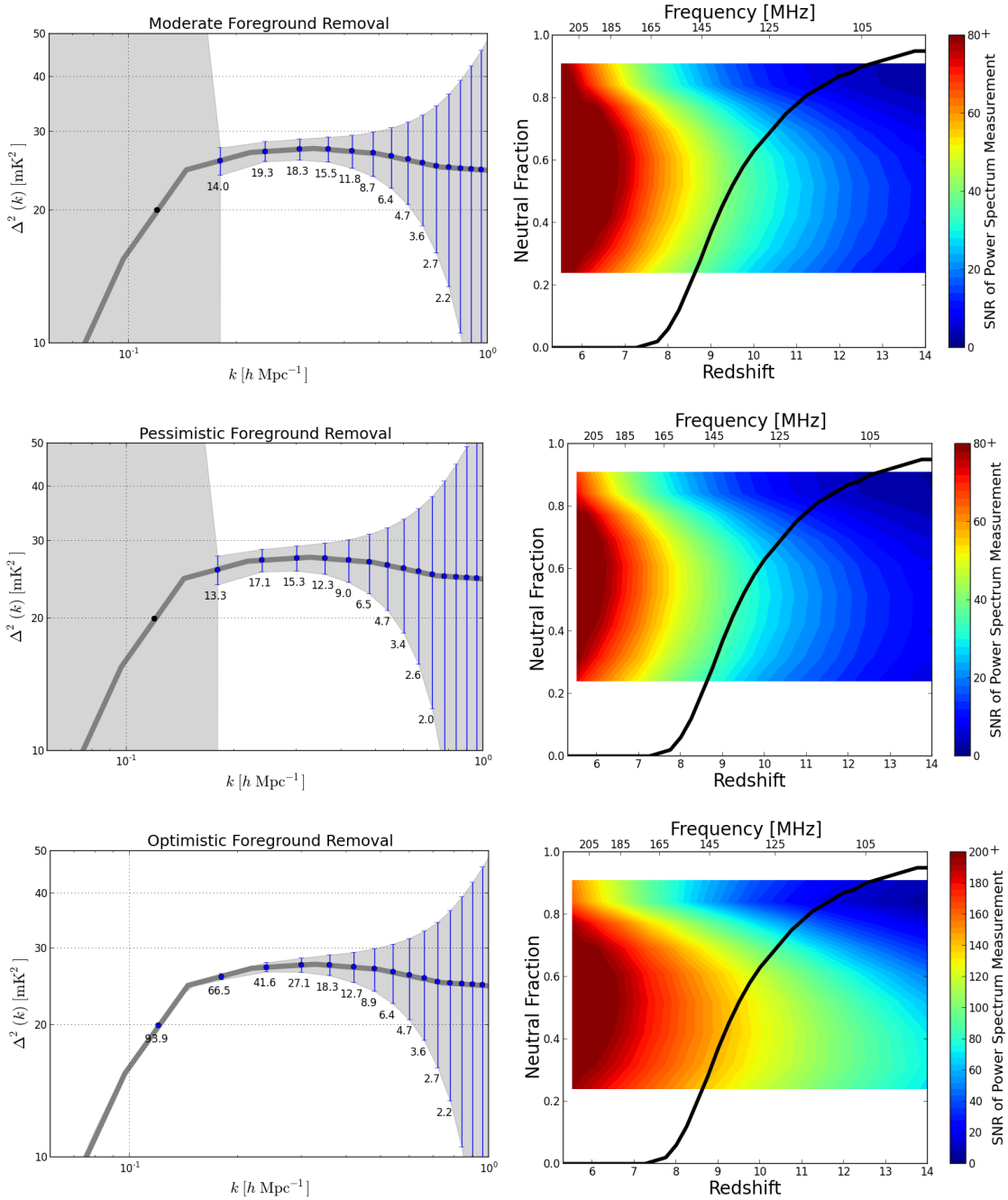
Figure 8-5: *Left:* Power spectrum constraints on the fiducial EoR model at $z = 9.5$ (53% ionization) for each of the three foreground removal models: moderate (top), pessimistic (middle) and optimistic (bottom). The shaded gray region shows the 1σ error range, whereas the location of the blue error bars indicate the binned sampling pattern; the binning is set by the bandwidth of 8 MHz. Black points without error bars indicate measurements allowed by instrumental parameters, but rendered unusable according to the foreground model. The net result of these measurements are 38σ, 32σ, and 133σ detections of the fiducial power spectrum for the moderate, pessimistic and optimistic models, respectively. Individual numbers below each error bar indicate the significance of the measurement in that bin. *Right:* Colored contours show the total SNR of a power spectrum detection as a function of redshift and neutral fraction for the three foreground removal models: moderate (top), pessimistic (middle) and optimistic (bottom). The black curve shows the fiducial evolution of the vanilla model; contour values off of the black curve are obtained by translating the fiducial model in redshift. This figure therefore allows one to examine the SNR for a far broader range of reionization histories than only those predicted by simulations with vanilla model parameters. Alternative evolution histories are less physically motivated, since a given set of EoR parameters does only predict one evolution history. The plotted sensitivities assume 8 MHz bandwidths are used to measure the power spectra, so not all points in the horizontal direction are independent. The incomplete coverage versus x_{HI} does not indicate that measurements cannot be made at these neutral fractions; rather, it is a feature of the `21cmFAST` code, and is explained in Section 8.3.1.

The left-hand panels of the figure show the constraints on the spherically averaged power spectrum at $z = 9.5$, the point of 50% ionization in this model, for the three foreground removal models. (The 50% ionization point generally corresponds to the peak power spectrum brightness at the scales of interest — as can be seen in Figure 8-4 — making its detection a key goal of reionization experiments [117, 22].) For the moderate model (top row), the errors amount to a 38σ detection of the 21 cm power spectrum at 50% ionization.

The right-hand panels of Figure 8.3.1.2 warrant somewhat detailed explanation. The three rows again correspond to the three foreground removal models. In each panel, the horizontal axis shows redshift and the vertical axis shows neutral fraction; thus this space spans a wide range of possible reionization histories. The black curve is the evolution of the vanilla model through this space. The colored contours show the signal-to-noise ratio of a HERA measurement of the 21 cm power spectrum at that point in redshift/neutral fraction space, where the power spectrum of a given

x_{HI} is extrapolated in redshift space as described in the beginning of Section 8.3.1. The colorscale is set to saturate at different values in each row: 80σ (moderate and pessimistic) and 200σ (optimistic). These sensitivities assume 8 MHz bandwidths are used to measure the power spectra, so not every value on the redshift-axis can be taken as an independent measurement. The non-uniform coverage versus ionization fraction (i.e. the white space at high and low values of x_{HI}) — which appears with different values in the panels of Figures 8.3.1.2, 8-6, 8-7, and 8-8 — is a feature of the `21cmFAST` code when attempting to produce power spectra for a set of input parameters at relatively even spaced values of ionization fraction. The black line is able to extend into the white region because it was generated to have uniform spacing in z as opposed to x_{HI}. The fact that these values are missing has minimal impact on the conclusions drawn in this work.

In the moderate model, the 50% ionization point of the fiducial power spectrum evolution is detected at $\sim 40\sigma$. However, we see that nearly every ionization fraction below $z \sim 9$ is detected with equally high significance. In general, the contours follow nearly vertical lines through this space. This implies that the evolution of sensitivity as a function of redshift (which is primarily driven by the system temperature) is much stronger than the evolution of power spectrum amplitude as a function of neutral fraction (which is primarily driven by reionization physics).

Figure 8-6 shows signal-to-noise contour plots for six different variations of our EoR parameters, using only the moderate foreground scenario. (The pessimistic and optimistic equivalents of this figure are shown in Figures 8-7 and 8-8, respectively.) In each panel, we have varied one parameter from the fiducial vanilla model. In particular, we choose the lowest and highest values of each parameter considered in Section 8.2.3. Since we extrapolate each power spectrum to a wide variety of redshifts, choosing only the minimum and maximum values leads to little loss of generality. Rather, we are picking extreme shapes and amplitudes for the power spectrum, and asking whether they can be detected if such a power spectrum were to correspond to a particular redshift. And, as with the vanilla model shown in Figure 8.3.1.2, it is clear that the moderate foreground removal scenario allows for the 21 cm power spectrum
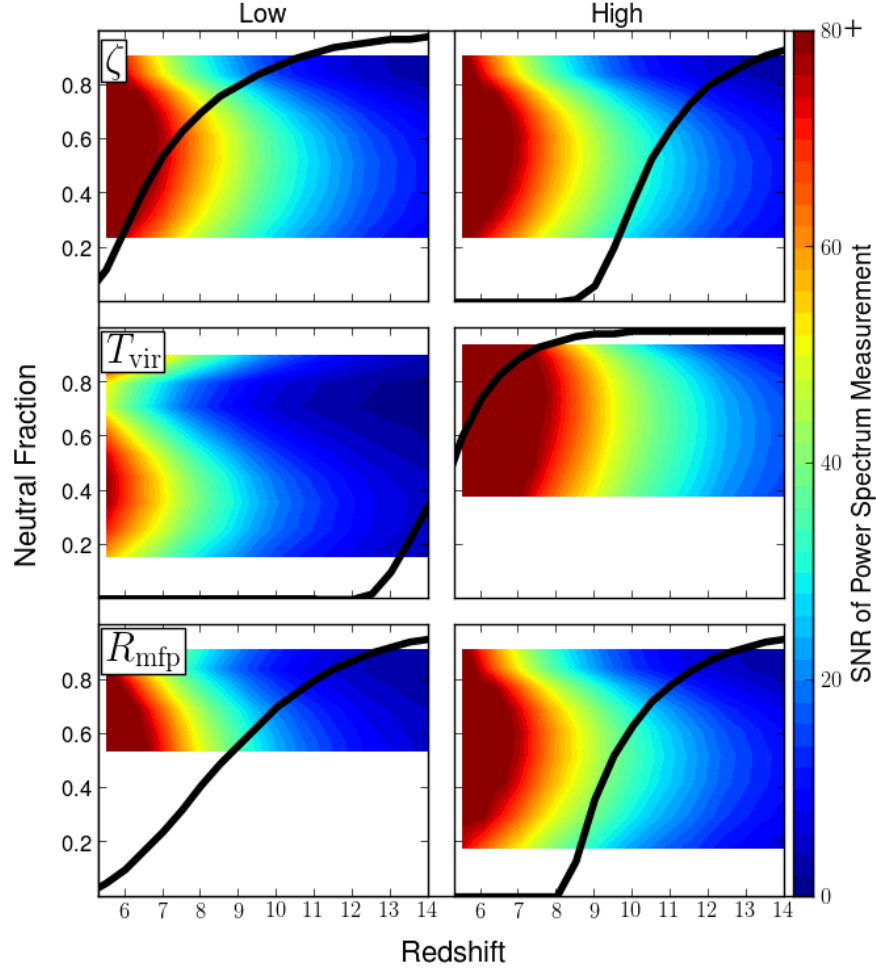
Figure 8-6: Signal-to-noise ratio of 21 cm power spectrum detections under the moderate foreground scenario for the high and low values of the parameters in our EoR models as functions of neutral fraction and redshift. In each panel, one parameter is varied, while the other two are held fixed at the "vanilla" values. The black curve shows the fiducial evolution for that set of model parameters. The incomplete coverage versus $x_{\rm HI}$ does not indicate that measurements cannot be made at these neutral fractions; rather, it is a feature of the `21cmFAST` code, and is explained in Section 8.3.1. *Top*: the ionizing efficiency, ζ. Values are $\zeta = 10$ (left) and $\zeta = 50$ (right). *Middle*: the minimum virial temperature of ionizing haloes, $T_{\rm vir}$. Values are $T_{\rm vir} = 1 \times 10^3$ K (left) and $T_{\rm vir} = 3 \times 10^5$ K (right). *Bottom*: the mean free path for ionizing photons through the IGM, $R_{\rm mfp}$. Values are $R_{\rm mfp} = 3$ Mpc (left) and $R_{\rm mfp} = 80$ Mpc (right). The moderate foreground removal scenario generically allows for a high significance measurement for nearly any reasonable reionization history.

426

to be detected with very high significance below $z \sim 8 - 10$, depending on the EoR model. Relative to the effects of system temperature, then, the actual brightness of the power spectrum as a function of neutral fraction plays a small role in determining the detectability of the cosmic signal. Of course, there is still a wide variety of power spectrum brightnesses; for a given EoR model, however, the relative power spectrum amplitude evolution as a function of redshift is fairly small.

There are also several more specific points about Figure 8-6 that warrant comment. Firstly, as stated in Section 8.2.3.2, the ionizing efficiency ζ has little effect on the shape of the power spectrum, but only on the timing and duration of reionization. This is clear from the identical sensitivity levels for both values of ζ, as well as for the vanilla model shown in Figure 8.3.1.2. Secondly, we reiterate that by tuning values of T_{vir} alone, we can produce ionization histories that are inconsistent with observations of the CMB and Lyman-α forest. In our analysis here, we extrapolate the power spectrum shapes produced by these extreme histories to more reasonable redshifts to show the widest range of possible scenarios. The fact that the fiducial evolution histories (black lines) of the T_{vir} row are wholly unreasonable is understood, and does not constitute an argument against this type of analysis.

8.3.1.3 Other Foreground Models

It is clear then, that the moderate foreground removal scenario will permit high sensitivity measurements of the 21 cm power with the next generation of experiments. Before considering what types of science these sensitivities will enable, it is worthwhile to consider the effects of the other foreground removal scenarios.

Our pessimistic scenario assumes — like the moderate scenario — that foregrounds irreparably corrupt $k_{\parallel}$ modes within the horizon limit plus 0.1 $h\mathrm{Mpc}^{-1}$, but also conservatively omits the coherent addition of partially redundant baselines in an effort to avoid multi-baseline systematics. As stated, this is the most conservative foreground case we consider. The achievable constraints on our fiducial vanilla power spectrum under this model were shown in the second row of Figure 8.3.1.2; Figure 8-7 shows the sensitivities for other EoR models. The sensitivity loss associated with
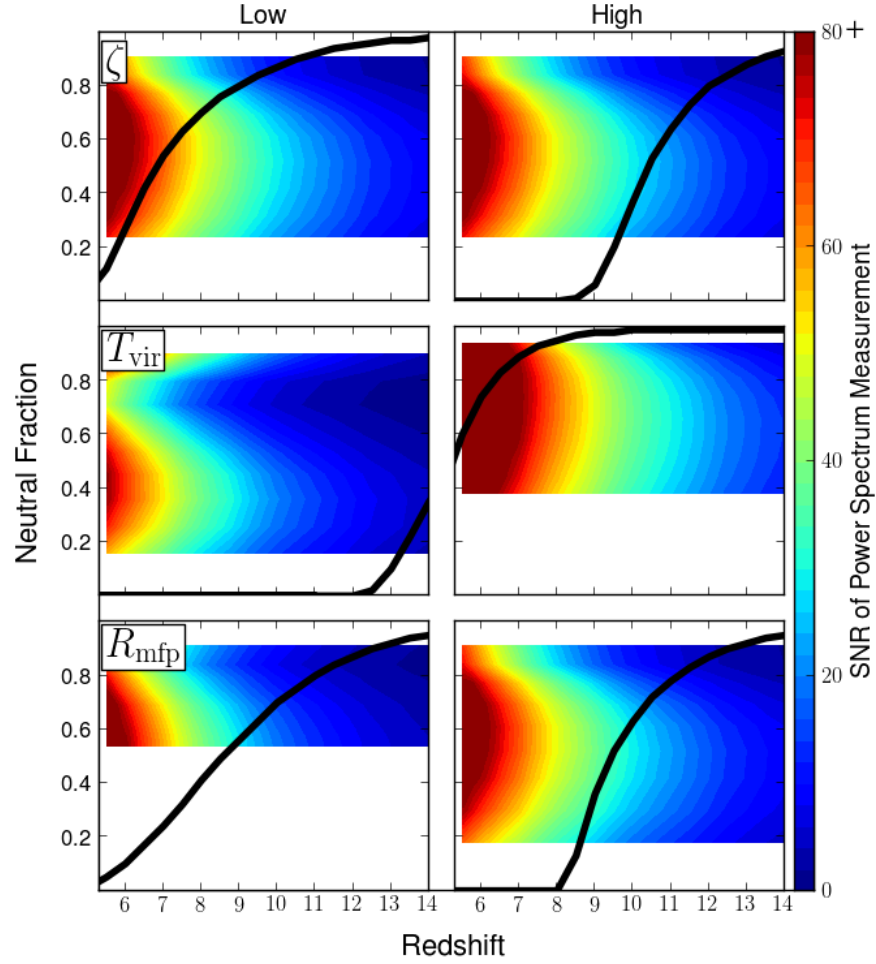
Figure 8-7: Same as Figure 8-6, but for the pessimistic foreground model. Note that the color-scale is the same as Figure 8-6 and saturates at 80σ.

coherently adding only instantaneously redundant baselines is fairly small, $\sim 20\%$. It should be noted that this pessimistic model affects only the thermal noise error bars relative to the moderate model; sample variance contributes the same amount of uncertainty in each bin. In an extreme sample variance limited case, the pessimistic and moderate models would yield the same power spectrum sensitivities. We will further explore the contribution of sample variance to these measurements in Section 8.3.3. Here we note that the pessimistic model generally increases the *thermal noise* uncertainties by 30–40% over the moderate model. This effect will be greater for an array with less instantaneous redundancy than the HERA concept design.

Finally, the sensitivity to the vanilla EoR model under the optimistic foreground removal scenario is shown in the bottom row of Figure 8.3.1.2. Figure 8-8 shows the sensitivity results for the other EoR scenarios. The sensitivities for the optimistic model are exceedingly high. Comparison of the top and bottom rows of Figure 8.3.1.2 shows that this model does not uniformly increase sensitivity across k-space, but rather the gains are entirely at low ks. This behavior is expected, since the effect of the optimistic model is to recover large scale modes that are treated as foreground contaminated in the other models. The sensitivity boosts come from the fact that thermal noise is very low at these large scales, since noise scales as k^3 while the cosmological signal remains relatively flat in $\Delta^2(k)$ space. We consider the effect of sample variance in these modes in Section 8.3.3.

8.3.2 The Timing and Duration of Reionization

One of the first key parameters that is expected from 21 cm measurement of the EoR power spectrum is the redshift at which the universe was 50% ionized, sometimes referred to as "peak reionization." The rationale behind this expectation is evident from Figure 8-4, where the power spectrum generically achieves peak brightness at $k \sim 0.1\ h\mathrm{Mpc}^{-1}$ for $x_{\mathrm{HI}} = 0.5$. However, given the steep increase of T_{sys}, one must ask if an experiment will truly have the sensitivity to distinguish the power spectrum at z_{peak} from those on either side. Figure 8-9 shows the error bars on our fiducial power spectrum model at 50% ionization ($z = 9.5$), as well as those on the neighboring
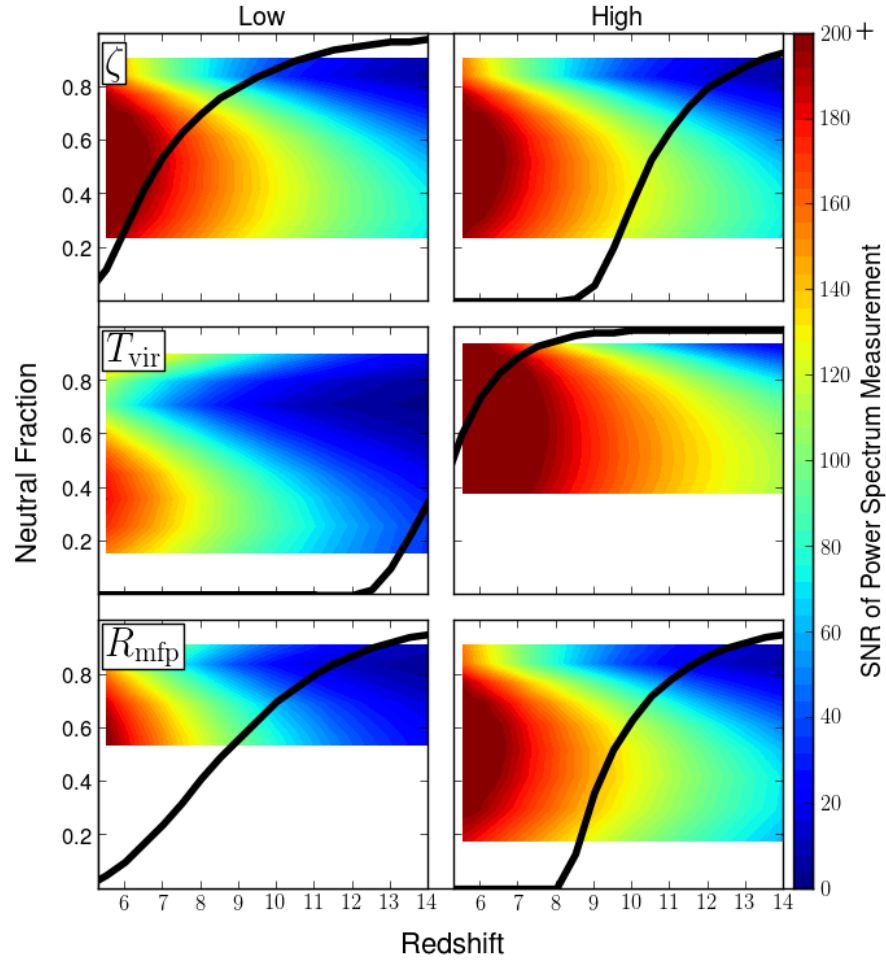
Figure 8-8: Same as Figures 8-6 and 8-7, but for the optimistic foreground model. Note that the color-scale has changed to saturate at 200σ.
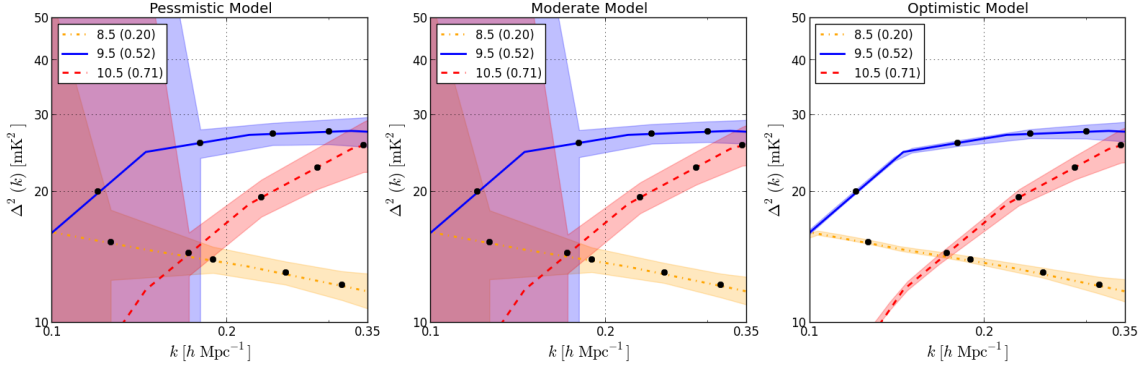
Figure 8-9: 1σ uncertainties in the measurements of the fiducial EoR power spectrum at redshifts 8.5, 9.5 and 10.5, corresponding to neutral fractions of 0.20, 0.52 and 0.71, respectively. Different panels show the results for the different foreground models: pessimistic (*left*), moderate (*middle*), and optimistic (*right*). The pessimistic and moderate scenarios should both permit measurements of $\Delta z \sim 1.0$. The optimistic scenario will allow for detailed characterization of the power spectrum evolution.

redshifts $z = 8.5$ and $z = 10.5$, under each of our three foreground models. In both the pessimistic and moderate models (left and middle panel), the $z = 8.5$ (20% neutral), $z = 9.5$ (52% neutral) and $z = 10.5$ (71% neutral) are distinguishable at the few sigma level. This analysis therefore suggests that it should be possible to identify peak reionization to within a $\Delta z \sim 1$, with a strong dependence on the actual redshift of reionization (since noise is significantly lower at lower redshifts).

It is worth noting, however, that even relatively high significance detections of the power spectrum ($\gtrsim 5$–10σ) may not permit one to distinguish power spectrum of peak reionization from those at nearby redshifts — especially as one looks to higher z. For our vanilla EoR model, we find a $\sim 10\sigma$ detection is necessary to distinguish the $z = 8.5$, 9.5, and 10.5 power spectra at the $> 1\sigma$ level. In fact, for this level of significance, nearly all of the power spectra at redshifts higher than peak reionization at $z = 9.5$ are indistinguishable given the steep rise in thermal noise. Even if the current generation of 21 cm telescopes does yield a detection of the 21 cm power spectrum, these first measurements do not guarantee stringent constraints on the peak redshift of reionization.

Finally, one can see that the high sensitivities permitted by the optimistic foreground model will allow a detailed characterization of the power spectrum amplitude
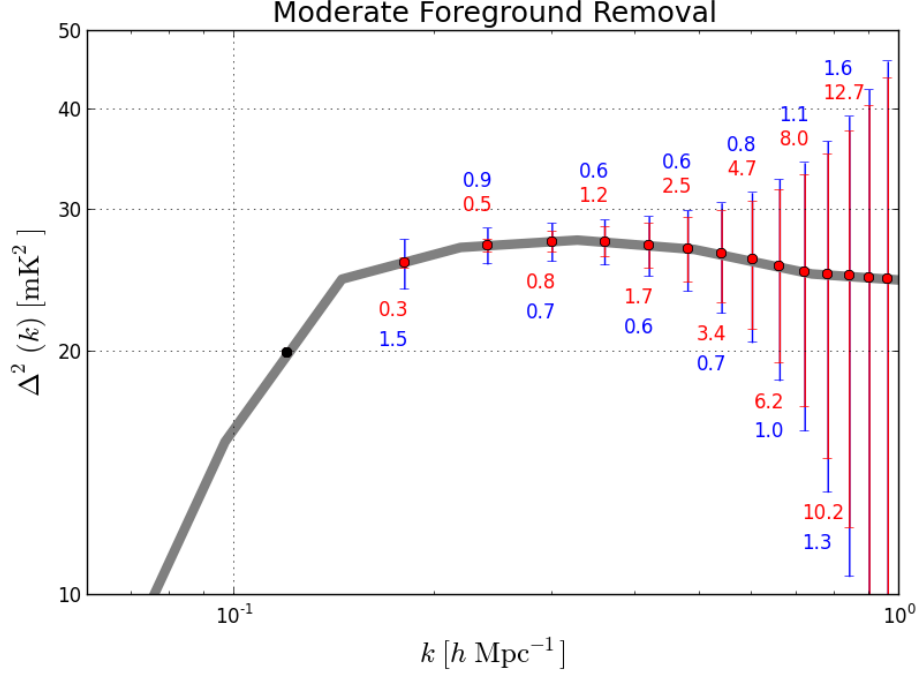
Figure 8-10: The breakdown of the error bars in the top-left panel of Figure 8.3.1.2 (vanilla EoR model and moderate foreground removal scenario). Red shows the contribution of thermal noise, blue the contribution of sample variance. The text shows the value of each contribution in mK^2 — *not* the significance of the detection, as in previous plots. Regions where the sample variance error dominates the thermal noise error are in the imaging regime. The placement of the numerical values above or below the error bar has no significance; it is only for clarity. Sample variance dominates the errors in the moderate foreground scenario on scales $k < 0.25$ $h\mathrm{Mpc}^{-1}$.

and slope as a function of redshift. We discuss exactly what kind of science this will enable (beyond detecting the timing and duration of reionization) in Section 8.3.5.

8.3.3   Sample Variance and Imaging

Given the high power spectrum sensitivities achievable under all of our foreground removal models, one must investigate the contributions of sample variance to the overall errors. For the moderate foreground model, Figure 8-10 shows the relative contribution of sample variance and thermal noise to the errors shown in the top-left panel of Figure 8.3.1.2. From this plot, it is clear that sample variance contributes over half of the total power spectrum uncertainty on scales $k < 0.3h\mathrm{Mpc}^{-1}$. If the power spectrum constituted the ultimate measurement of reionization, this would be

an argument for a survey covering a larger area of sky. For our HERA concept array, which drift-scans, this is not possible, but may be for phased-array designs. However, the sample variance dominated regime is very nearly equivalent to the imaging regime: thermal noise is reduced to the point where individual modes have an SNR $\gtrsim$ 1. Therefore, using a filter to remove the wedge region (e.g. Pober et al. [182]), a collecting area of 0.1 km^2 should provide sufficient sensitivity to image the Epoch of Reionization over $\sim$ 800 sq. deg. (6 hours of right ascension $\times$ 8.7° FHWM) on scales of 0.1–0.25 hMpc^{-1}. We note that the HERA concept design is not necessarily optimized for imaging; other configurations may be better suited if imaging the EoR is the primary science goal.

The effect of the other foreground models on imaging is relatively small. The poorer sensitivities of the pessimistic model push up thermal noise, lowering the highest k that can be imaged to $k \sim 0.2 h$Mpc^{-1}. The optimistic foreground model recovers significant SNR on the largest scales, to the point where sample variance dominates all scales up to 0.3 hMpc^{-1}. The effects of foregrounds and the wedge on imaging with a HERA-like array will be explored in future work.

8.3.4 Characteristic Features of EoR Power Spectrum

Past literature has discussed two simple features of the 21 cm power spectra to help distinguish between models: the slope of the power spectrum and the sharp drop in power (the "knee") on the largest scales [139]. In particular, the mass of the halos driving reionization (parametrized in this analysis by the minimum virial temperature of ionizing halos) should affect the slope of the power spectrum. Since more massive halos are more highly clustered, they should concentrate power on larger scales, yielding a flatter slope. The second row of Figure 8-4 suggests this effect is small, although not implausible. The knee of the power spectrum at large scales should correspond to the largest ionized bubble size, since there will be little power on scales larger than these bubbles [73]. The position of the knee should be highly sensitive to the mean free path for ionizing photons through the IGM, since this sets how large bubbles can grow. This argument is indeed confirmed by the third row of Figure 8-4, where the
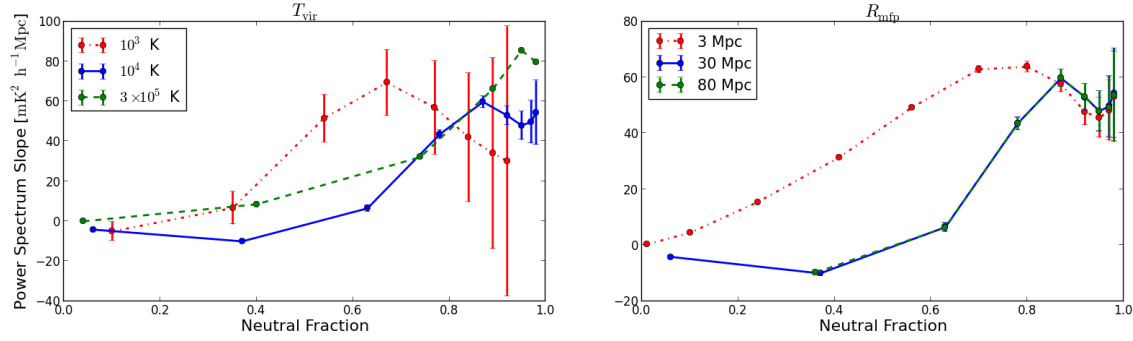
Figure 8-11: The power spectrum slope in units of $mK^2 h^{-1}Mpc$ between $k = 0.1$ and $1.0 h Mpc^{-1}$ as a function of neutral fraction for various EoR models. Note that error bars are plotted on all points and correspond to the redshift of a given neutral fraction for that model. *Left*: Different values of T_{vir}. *Right*: Different values of R_{mfp}. While different values of T_{vir} and R_{mfp} produce considerable changes in the power spectrum slope, it will be difficult to unambiguously interpret its physical significance.

smaller values of R_{mfp} lack significant power on large scales compared to those models with larger values. Unfortunately, since our third parameter ζ does not change the shape of the power spectrum, constraining different values of ζ will not be possible with only a shape-based analysis. In this section we first extend these qualitative arguments based on salient features in the power spectra, and then present a more quantitative analysis on distinguishing models in Section 8.3.5.

To quantify the slope of the power spectrum, we fit a straight line to the predicted power spectrum values between $k = 0.1$–1.0 $h Mpc^{-1}$. When we refer to measuring the slope, we refer to measuring the slope of this line, given the error bars in the k-bins between 0.1–1.0 $h Mpc^{-1}$. This choice of fit is not designed to encompass the full range of information contained in measurements of the power spectrum shape. Rather, the goal of this section is to find simple features of the power spectrum that can potentially teach us about the EoR without resorting to more sophisticated modeling.

Figure 8-11 shows the evolution of the slope of the linear fit to the power spectrum over the range $k = 0.1$–1.0 $h Mpc^{-1}$, as a function of neutral fraction for several EoR models. Error bars in both panels correspond to the error measured under the moderate foreground model for a given neutral fraction in the fiducial history

of a model. This means that, e.g., the high neutral fractions in the $T_{\mathrm{vir}} = 10^3$ K curve have thermal noise errors corresponding to $z \sim 20$, outside the range of many proposed experiments. Given that caveat, it does appear that the low-mass ionizing galaxies produce power spectra with significantly steeper slopes at moderate neutral fractions than those models where only high mass galaxies produce ionizing photons. However, it is also clear that small-bubble (i.e. low mean free path) models can yield steep slopes. Therefore, while there may be some physical insight to be gleaned from measuring the slope of the power spectrum and its evolution, even the higher sensitivity measurements permitted by the optimistic foreground model may not be enough to break these degeneracies. In Section 8.3.5, we specifically focus on the kind of information necessary to disentangle these effects.

A comparison of the error bars in the moderate and optimistic foreground scenario measurements of the vanilla power spectrum (rows one and three in Figure 8.3.1.2) reveals the difficulty in recovering the position of the knee without foreground subtraction: foreground contamination predominantly affects low k modes, rendering large scale features like the knee inaccessible. In particular, the additive component of the horizon wedge severely restricts the large scale information available to the array. Without probing large scales, confirming the presence (or absence) of a knee feature is likely to be impossible. However, Figure 8.3.1.2 does show that if foreground removal allows for the recovery of these modes, the knee can be detected with very high significance, even the presence of sample variance.

8.3.5 Quantitative Constraints on Model Parameters

In previous sections, we considered rather large changes to the input parameters of the `21cmFAST` model. These gave rise to theoretical power spectra that exhibited large qualitative variations, and encouragingly, we saw that such variations should be easily detectable using next-generation instruments such as HERA. We now turn to what would be a natural next step in data analysis following a broad-brush, qualitative discrimination: a determination of best-fit values for astrophysical parameters. In this section, we forecast the accuracy with which T_{vir}, R_{mfp}, and ζ can be measured

by a HERA-like instrument, paying special attention to degeneracies. In many of the plots, we will omit the pessimistic foreground scenario, for often the results from it are identical to (and visually indistinguishable from) those from moderate foregrounds. Ultimately, our results will suggest parameter constraints that are smaller than one can justifiably expect given the reasonable, but non-negligible uncertainty surrounding simulations of reionization [245]. Our final error bar predictions (which can be found in Table 8.3) should therefore be interpreted cautiously, but we do expect the qualitative trends in our analysis to continue to hold as theoretical models improve.

8.3.6 Fisher matrix formalism for errors on model parameters

To make our forecasts, we use the Fisher information matrix $\mathbf{F}$, which takes the form

$$\mathbf{F}_{ij} \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle = \sum_{k,z} \frac{1}{\varepsilon^2(k,z)} \frac{\partial \Delta^2(k,z)}{\partial \theta_i} \frac{\partial \Delta^2(k,z)}{\partial \theta_j}, \qquad (8.3)$$

where $\mathcal{L}$ is the likelihood function (i.e. the probability distribution for the measured data as a function of model parameters), $\varepsilon(k,z)$ is the error on $\Delta^2(k,z)$ measurements as a function of wavenumber k and redshift z, and $\boldsymbol{\theta} = (T_{\mathrm{vir}}/T_{\mathrm{vir}}^{\mathrm{fid}}, R_{\mathrm{mfp}}/R_{\mathrm{mfp}}^{\mathrm{fid}}, \zeta/\zeta^{\mathrm{fid}})$ is a vector of the parameters that we wish to measure, divided by their fiducial values[8]. The second equality in Equation (8.3) follows from assuming Gaussian errors, and picking k-space and redshift bins in a way that ensures that different bins are statistically independent [212], as we have done throughout this paper. Implicit in our notation is the understanding that all expectation values and partial derivatives are evaluated at fiducial parameter values. Having computed the Fisher matrix, one can obtain the error bars on the ith parameter by computing $1/\sqrt{\mathbf{F}_{ii}}$ (when all other parameters are known already) or $(\mathbf{F}^{-1})_{ii}$ (when all parameters are jointly estimated from the data). The Fisher matrix thus serves as a useful guide to the error properties of a measurement, albeit one that has been performed optimally. Moreover, because Fisher information is additive (as demonstrated explicitly in Equation [8.3]), one can

---

[8]Scaling out the fiducial values of course represents no loss of generality, and is done purely for numerical convenience.
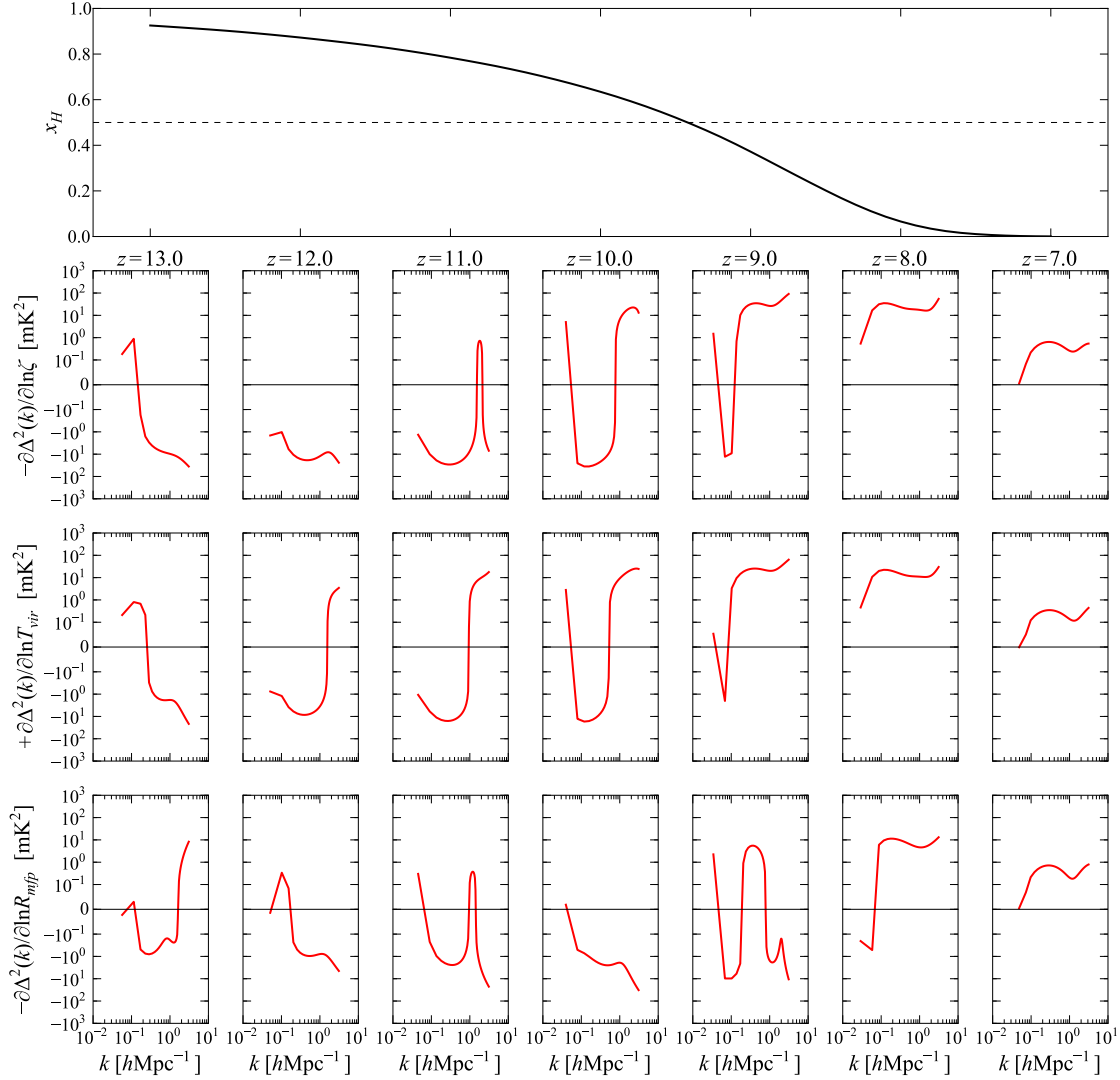
Figure 8-12: Power spectrum derivatives as a function of wavenumber k and redshift z. Each of the lower three rows shows derivatives with respect to a different parameter in our three-parameter model, and the top panel (aligned in redshift with the bottom panels) shows the corresponding neutral fraction. Because our parameter vector $\boldsymbol{\theta}$ (Equation [8.3]) contains non-dimensionalized parameters, the derivatives $\partial \Delta^2(k,z)/\partial \theta_i$ are equivalent (if evaluated at the fiducial parameter values) to the logarithmic derivatives shown here. Note that the derivatives with respect to ζ and R_{mfp} have been multiplied by -1 to facilitate later comparisons. The vertical axes for the derivatives are linear between -10^{-1} and 10^1, and are logarithmic outside that range. From this figure, we see that while the lowest redshifts are easy to access observationally, the model parameters are highly degenerate. The higher redshifts are less degenerate, but thermal noise and foregrounds make observations difficult.

437

conveniently examine which wavenumbers and redshifts contribute the most to the parameter constraints, and we will do so later in the section.

From Equation (8.3), we see that it is the derivatives of the power spectrum with respect to the parameters that provide the crucial link between measurement and theory. If a large change in the power spectrum results from a small change in a parameter — if the amplitude of a power spectrum derivative is large — a measurement of the power spectrum would clearly place strong constraints on the parameter in question. This is a property that is manifest in $\mathbf{F}$. Also important are the shapes of the power spectrum derivatives in (k, z) space. If two power spectrum derivatives have similar shapes, changes in one parameter can be mostly compensated for by a change in the second parameter, leading to a large degeneracy between the two parameters. Mathematically, the power spectrum derivatives can be geometrically interpreted as vectors in (k, z) space, and each element of the Fisher matrix is a weighted dot product between a pair of such vectors [212]. Explicitly, $\mathbf{F}_{ij} = \mathbf{w}_i \cdot \mathbf{w}_j$, where

$$\mathbf{w}_i(k, z) \equiv \frac{1}{\varepsilon(k, z)} \frac{\partial \Delta^2(k, z)}{\partial \theta_i}, \tag{8.4}$$

with the different elements of the vector corresponding to different values of k and z. If two $\mathbf{w}$ vectors have a large dot product (i.e. similar shapes), the Fisher matrix will be near-singular, and the joint parameter constraints given by $\mathbf{F}^{-1}$ will be poor.

8.3.6.1 Single-Redshift Constraints

We begin by examining how well each reionization parameter can be constrained by observations at several redshifts spanning our fiducial reionization model. In Figure 8-12, we show some example power spectrum derivatives[9] as a function of k and z. Note

---

[9]Because `21cmFAST` produces output at k-values that differ from those naturally arising from our sensitivity calculations, it was necessary to interpolate the outputs when computing the derivatives (which were approximated using finite-difference methods). For this paper, we chose to fit the `21cmFAST` power spectra to sixth-order polynomials in $\ln k$, finding such a scheme to be a good balance between capturing all the essential features of the power spectrum derivatives while not over-fitting any "noise" in the theoretical simulations. Alternate approaches such as performing cubic splines, or fitting to fifth- or seventh-order polynomials were tested, and do not change our results in any meaningful ways. Finally, to safeguard against generating numerical derivatives that are dominated by the intrinsic numerical errors of the simulations, we took care to choose finite-

that the last two rows of the figure show the *negative* derivatives. For reference, the top panel of the figure shows the corresponding evolution of the neutral fraction. At the lowest redshifts, the power spectrum derivatives essentially have the same shape as the power spectrum itself. To understand why this is so, note that at late times, a small perturbation in parameter values mostly shifts the time at which reionization reaches completion. As reionization nears completion, the power spectrum decreases proportionally in amplitude at all k due to a reduction in the overall neutral fraction, so a parameter shift simply causes an overall amplitude shift. The power spectrum derivatives are therefore roughly proportional to the power spectrum. In contrast, at high redshifts the derivatives have more complicated shapes, since changes in the parameters affect the detailed properties of the ionization field.

Importantly, we emphasize that for parameter estimation, the "sweet spot" in redshift can be somewhat different from that for a mere detection. As mentioned in earlier sections, the half-neutral point of $x_H = 0.5$ is often considered the most promising for a detection, since most theoretical calculations yield peak power spectrum brightness there. This "detection sweet spot" may shift slightly towards lower redshifts because thermal noise and foregrounds decrease with increasing frequency, but is still expected to be close to the half-neutral point. For parameter estimation, however, the most informative redshifts may be significantly lower. Consider, for instance, the signal at $z = 8$, where $x_H = 0.066$ in our fiducial model. Figure 8-3 reveals that the power spectrum here is an order of magnitude dimmer than at $z = 9.5$ or $z = 9$, where $x_H = 0.5$ and $x_H = 0.37$ respectively. However, from Figure 8-12 we see that the power spectrum derivatives at $z = 8$ are comparable in amplitude to those at higher redshifts/neutral fraction. Intuitively, at $z = 8$ the dim power spectrum is compensated for by its rapid evolution (due to the sharp fall in power towards the end of reionization). Small perturbations in model parameters and the resultant changes in the timing of the end of reionization therefore cause large changes in the theoretical power spectrum. There is thus a large information content in a $z = 8$ power spectrum measurement.

---

difference step sizes that were not overly fine.
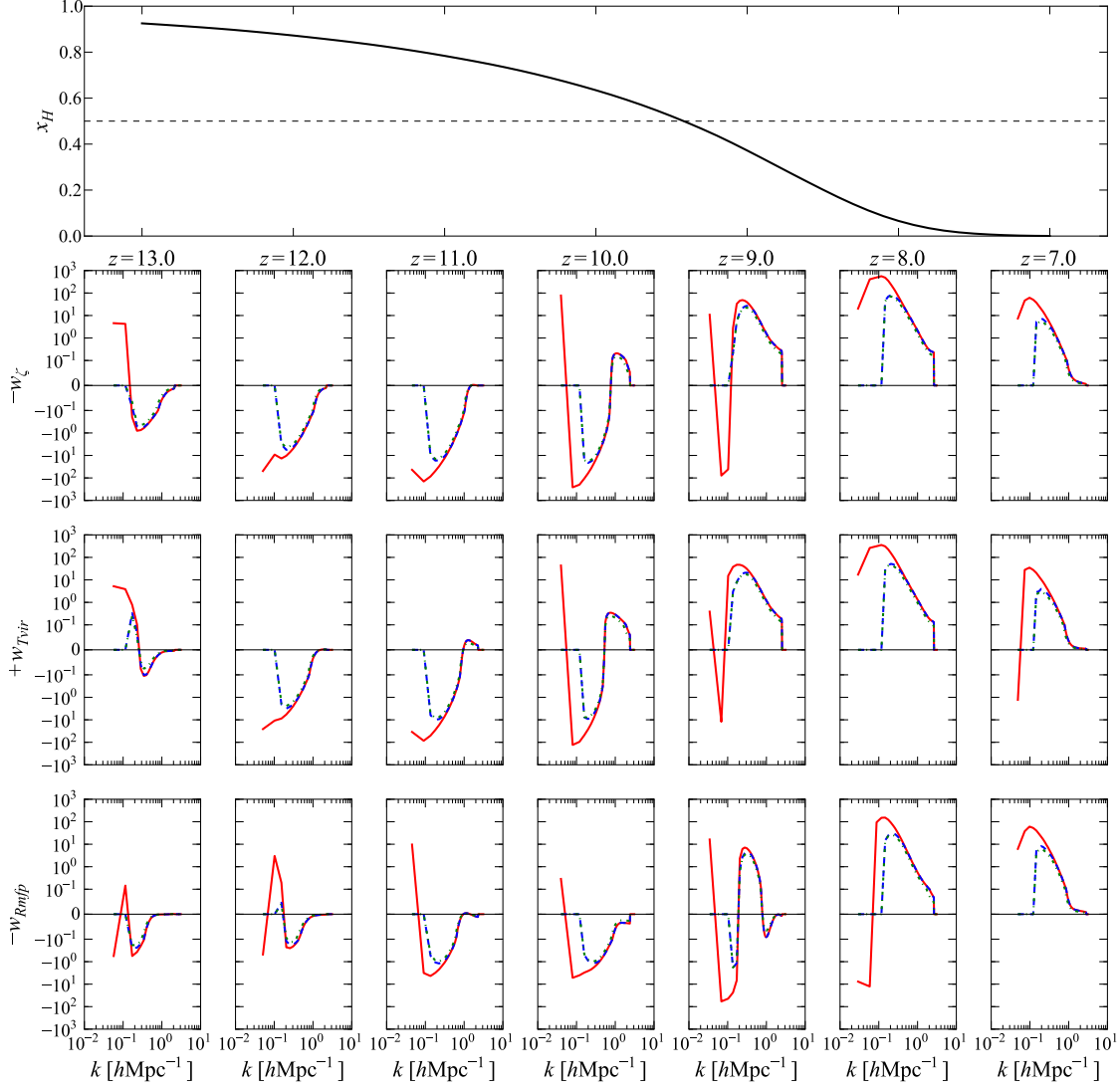
Figure 8-13: Similar to Figure 8-12, but inversely weighted by the error on the power spectrum measurement, i.e. plots of $\mathbf{w}_i$ from Equation (8.4). These weighted derivatives are computed for HERA for the optimistic (solid red curves), moderate (dashed blue curves), and pessimistic (dot-dashed green curves) foreground models. The pessimistic curves are essentially indistinguishable from the moderate curves. The top panel shows the corresponding evolution of the neutral fraction. Just as with Figure 8-12, the vertical axes are linear from -10^{-1} to 10^1 and logarithmic elsewhere, and $\mathbf{w}_{\mathrm{Rmfp}}$ and $\mathbf{w}_\zeta$ have been multiplied by -1 to facilitate comparison with $\mathbf{w}_{\mathrm{Tvir}}$. With foregrounds and thermal noise, power spectrum measurements become difficult at low and high k values, and constraints on model parameters become more degenerate.

440

When thermal noise and foregrounds are taken into account, a $z = 8$ measurement becomes even more valuable for parameter constraints than those at higher redshifts/neutral fractions. This can be seen in Figure 8-13, where we weight the power spectrum derivatives by the inverse measurement error[10] for HERA, producing the quantity $\mathbf{w}_i$ as defined in Equation (8.4). In solid red are the weighted derivatives for the optimistic foreground model, while the dashed blue curves are for the moderate foreground model. The pessimistic case is shown using dot-dashed green curves, but these curves are barely visible because they are essentially indistinguishable from those for the moderate foregrounds. In all cases, the derivatives peak — and therefore contribute the most information — at $z = 8$. Squaring and summing these curves over k, one can compute the diagonal elements of the Fisher matrix on a per-redshift basis. Taking the reciprocal square root of these elements gives the error bars on each parameter assuming (unrealistically) that all other parameters are known. The results are shown in Figure 8-14. For all three parameters, these single-parameter, per-redshift fits give the best errors at $z = 8$. At $z = 7$ the neutral fraction is simply too low for there to be any appreciable signal (with even the rapid evolution unable to sufficiently compensate), and at higher redshifts, thermal noise and foregrounds become more of an influence.

Of course, what is not captured by Figure 8-14 is the reality that one must fit for all parameters simultaneously (since none of our three parameters are currently strongly constrained by other observational probes). In general, our ability to constrain model parameters is determined not just by the amplitudes of the power spectrum derivatives and our instrument's sensitivity, but also by parameter degeneracies. As an example, notice that at $z = 7$, all the power spectrum derivatives shown in Figure 8-12 have essentially identical shapes up to a sign. This means that shifts in one parameter can be (almost) perfectly nullified by a shift in a different parameter; the parameters are degenerate. These degeneracies are inherent to the theoretical model, since they are clearly visible even in Figure 8-12, where the power spectrum derivatives are shown

---

[10]Whereas in previous sections the power spectrum sensitivities were always computed assuming a bandwidth of 8 MHz, in this section we vary the bandwidth with redshift so that a measurement centered at redshift z uses all information from $z \pm 0.5$.
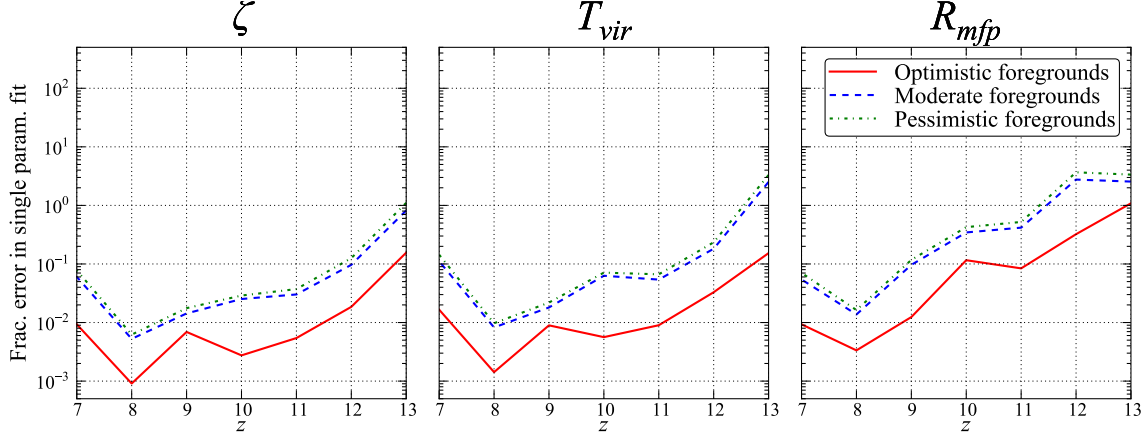
Figure 8-14: Fractional 1σ errors (1σ errors divided by fiducial values) as a function of redshift, with measurements at each redshift fit independently. The errors on each parameter assume (unrealistically) that all other parameters are already known in the fit. Solid red curves give optimistic foreground model predictions; dashed blue curves give moderate foreground model predictions; dot-dashed green curves give pessimistic foreground model predictions. In all models, and for all three parameters, the best errors are obtained at $z = 8$. At $z = 7$, the power spectrum has too small of an amplitude to yield good signal-to-noise, and at higher redshifts thermal noise poses a serious problem.

without the instrumental weighting. With this in mind, we see that even though Figure 8-14 predicts that observing the power spectrum at $z = 7$ alone would give reasonable errors if there were somehow no degeneracies (making a single parameter fit the same as a simultaneous fit), such a measurement would be unlikely to yield any useful parameter constraints in practice. To only a slightly lesser extent, the same is true for $z = 8$, where the degeneracy between R_{mfp} and the other parameters is broken slightly, but T_{vir} and ζ remain almost perfectly degenerate.

The situation becomes even worse when one realizes that measurements at low and high k are difficult due to foregrounds and thermal noise, respectively. Many of the distinguishing features between the curves in Figure 8-12 were located at the extremes of the k axis, and from Figure 8-13, we see that such features are obliterated by an instrumental sensitivity weighting (particularly for the pessimistic/moderate foregrounds). This increases the level of degeneracy. As an aside, note that with the lowest and highest k values cut out, the bulk of the information originates from $k \sim 0.05\ h\mathrm{Mpc}^{-1}$ to $\sim 1\ h\mathrm{Mpc}^{-1}$ for the optimistic model and $k \sim 0.2\ h\mathrm{Mpc}^{-1}$ to $\sim$

442
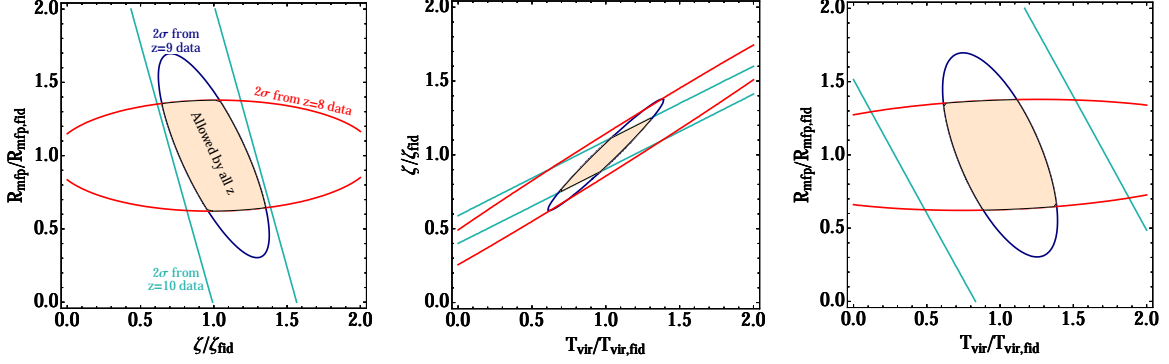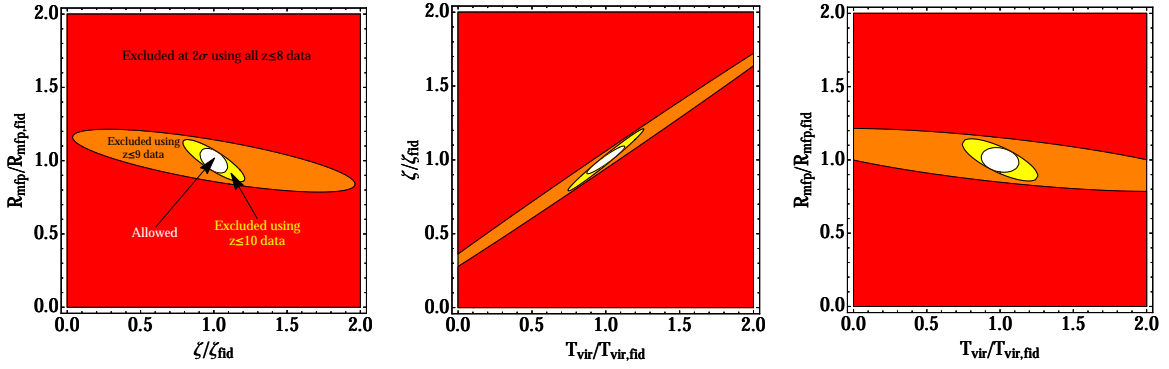
Figure 8-15: Pairwise parameter constraints for the moderate foreground model, shown as 2σ exclusion regions. For each pair of parameters, the third parameter has been marginalized over. The top row shows 2σ constraints from each redshift when fit independently; the light orange regions are not ruled out by data from any redshifts. The bottom row shows the constraints from a joint fit over multiple redshifts. Each color represents a portion of parameter space that can be *excluded* by including data up to a certain redshift. The white "allowed" region represents the final constraints from including all measured redshifts. In both cases, a $z = 7$ measurement alone does not provide any non-trivial constraints, but helps with degeneracy-breaking in the joint redshift fits (bottom row). As one moves to higher and higher redshifts, power spectrum measurements probe different astrophysical processes, resulting in a shift in the principal directions of the exclusion regions. Including higher redshifts tightens parameter constraints, but no longer helps beyond $z = 10$ due to increasing thermal noise.

1 $h\mathrm{Mpc}^{-1}$ for the pessimistic and moderate models. (Recall again that since elements of the Fisher matrix are obtained by taking pairwise products of the rows of Figure 8-13 and summing over k and z, the square of each weighted power spectrum derivative curve provides a rough estimate for where information comes from.) Matching these ranges to the fiducial power spectra in Figure 8-3 confirms the qualitative discussion presented in Section 8.3.4, where we saw that slope of the power spectrum from $k \sim 0.1\ h\mathrm{Mpc}^{-1}$ to $\sim 1\ h\mathrm{Mpc}^{-1}$ is potentially a useful source of information regardless of foreground scenario, but that the "knee" feature at $k \lesssim 0.1\ h\mathrm{Mpc}^{-1}$ will likely only be accessible with optimistic foregrounds. This is somewhat unfortunate, for a comparison of Figures 8-12 and 8-13 reveals that measurements at low and high k would potentially be powerful breakers of degeneracy, were they observationally feasible.

8.3.6.2 Breaking degeneracies with multi-redshift observations

Absent a situation where the lowest and highest k values can be probed, the only way to break the serious degeneracies in the high signal-to-noise measurements at $z = 7$ and $z = 8$ is to include higher redshifts, even though thermal noise and foreground limitations dictate that such measurements will be less sensitive. Higher redshift measurements break degeneracies in two ways. First, one can see that at higher redshifts, the power spectrum derivatives have shapes that are both more complicated and less similar, thanks to non-trivial astrophysics during early to mid-reionization. Second, a joint fit over multiple redshifts can alleviate degeneracies even if the parameters are perfectly degenerate with each other at every redshift when fit on a per-redshift basis. Consider, for example, the weighted power spectrum derivatives for the moderate foreground model in Figure 8-13. For both $z = 7$ and $z = 8$, the derivatives for all three parameters are identical in shape; at both redshifts, any shift in the best-fit value of a parameter can be compensated for by an appropriate shift in the other parameters without compromising the goodness-of-fit. At $z = 8$, for instance, a given fractional increase in ζ can be compensated for by a slightly larger decrease in R_{mfp}, since $w_{R_{\mathrm{mfp}}}$ has a slightly larger amplitude than w_ζ. However, this

only works if the redshifts are treated independently. If the data from $z = 7$ and $z = 8$ are jointly fit, the aforementioned parameter shifts would result in a worse overall fit, because w_{Rmfp} and w_ζ have roughly equal amplitudes at $z = 7$, demanding fractionally equal shifts. In other words, we see that because the *ratios* of different weighted parameter derivatives are redshift-dependent quantities, joint-redshift fits can break degeneracies even when the parameters would be degenerate if different redshifts were treated independently. It is therefore crucial to make observations at a wide variety of redshifts, and not just at the lowest ones, where the measurements are easiest.

To see how degeneracies are broken by using information from multiple redshifts, imagine a thought experiment where one began with measurements at the lowest (least noisy) redshifts, and gradually added higher redshift information, one redshift at a time. Figures 8-15 and 8-16 show the results for the moderate and optimistic foreground scenarios respectively. (Here we omit the equivalent figure for the pessimistic model completely, because the results are again qualitatively similar to those for the moderate model.) In each figure are 2σ constraints for pairs of parameters, having marginalized over the third parameter by assuming that the likelihood function is Gaussian (so that the covariance of the measured parameters is given by the inverse of the Fisher matrix). One sees that as higher and higher redshifts are included, the principal directions of the exclusion ellipses change, reflecting the first degeneracy-breaking effect highlighted above, namely, the inclusion of different, more-complex and less-degenerate astrophysics at higher redshifts. To see the second degeneracy-breaking effect, where per-redshift degeneracies are broken by joint redshift fits, we include in both figures the constraints that arise after combining results from redshift-by-redshift fits (shown as contours for each redshift), as well as the constraints from fitting multiple redshifts simultaneously (shown as cumulative exclusion regions).

For the moderate foreground scenario, we find that non-trivial constraints cannot be placed using $z = 7$ data alone, hence the omission of a $z \leq 7$ exclusion region from Figure 8-15. However, we note that the constraints using $z \leq 8$ data (red contours/exclusion regions) are substantially tighter in the bottom panel than in
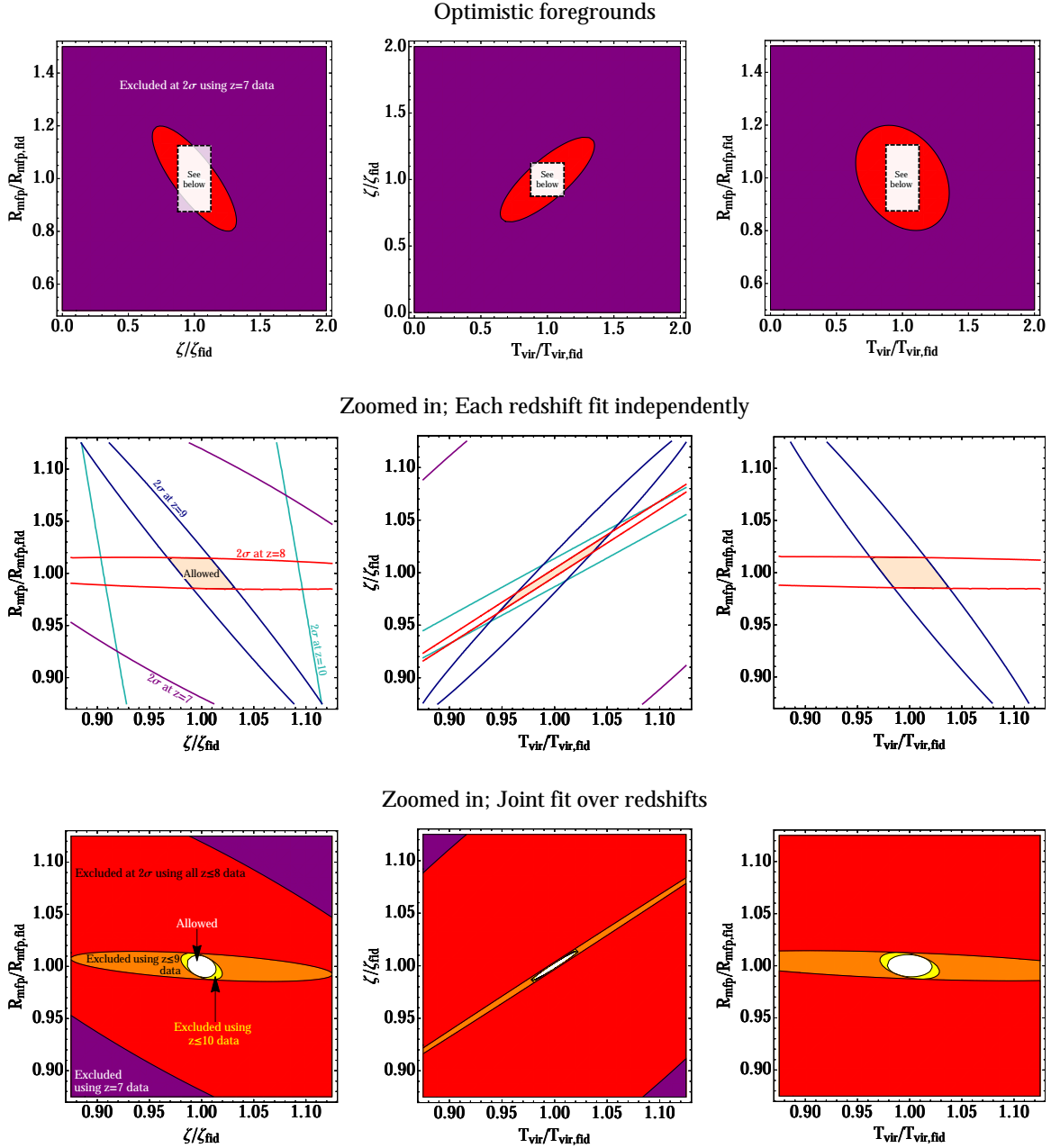
Figure 8-16: Similar to Figure 8-15, but for the optimistic foreground model. The top row shows the exclusion region from using $z = 7$ data alone. The middle and bottom rows show zoomed-in parameter space plots for the redshift-by-redshift and simultaneous fits respectively. (Since $z = 7$ is the lowest redshift in our model, the top panel is the same for both types of fit). The constraints in this optimistic foreground scenario are seen to be better than those predicted for the moderate foreground model by about a factor of four.

the top panel of the figure. This means that the $z = 7$ power spectrum can break degeneracies in a joint fit, even if the constraints from it alone are too degenerate to be useful. A similar situation is seen to be true for a $z = 10$ measurement, which is limited not just by degeneracy, but also by the higher thermal noise at lower frequencies. Except for the ζ-$T_{\rm vir}$ parameter space, adding $z = 10$ information in an independent fashion does not further tighten the constraints beyond those provided by $z \leq 9$. But again, when a joint fit (bottom panel of Figure 8-15) is performed, this information is useful even though it was noisy and degenerate on its own. We caution, however, that this trend does *not* persist beyond $z = 10$, in that $z \geq 11$ measurements are so thermal-noise dominated that their inclusion has no effect on the final constraints. Indeed, the "allowed" regions in both Figures 8-15 and 8-16 include all redshifts, but are visually indistinguishable from ones calculated without $z \geq 11$ information.[11]
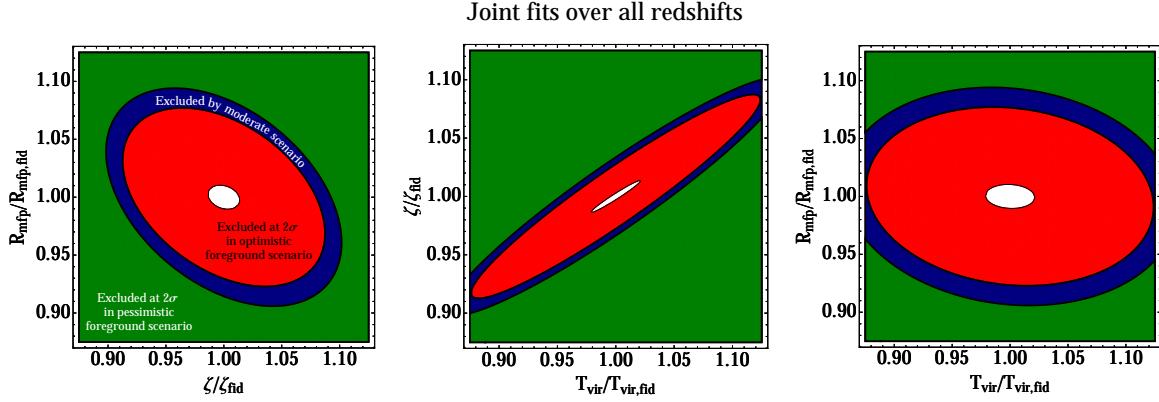
Figure 8-17: A comparison between the 2σ exclusion regions for pessimistic (green), moderate (blue), and optimistic (red) foregrounds, assuming that power spectra at all measured redshifts are fit simultaneously. Going from the pessimistic foreground model to the moderate model gives only marginal improvement; going from the moderate to the optimistic model reduces errors from the 5% level to the 1% level.

Comparing the predictions for the moderate foreground model to those of the optimistic foreground model (Figure 8-16), several differences are immediately apparent. Whereas the $z = 7$ power spectrum alone could not place non-trivial parameter

---

[11]We emphasize that in our analysis we have only considered the reionization epoch. Thus, while we find that observations of power spectra at $z \geq 11$ do not add very much to measurements of reionization parameters like $T_{\rm vir}$, ζ, and $R_{\rm mfp}$, they are expected to be extremely important for constraining X-ray physics prior to reionization, as discussed in [147] and [42].

Foreground Model	$\Delta T_\mathrm{vir}/T_\mathrm{vir,fid}$	$\Delta\zeta/\zeta_\mathrm{fid}$	$\Delta R_\mathrm{mfp}/R_\mathrm{mfp,fid}$
Moderate	0.062	0.044	0.039
Pessimistic	0.071	0.051	0.047
Optimistic	0.011	0.0069	0.0052

Table 8.3: Reionization Parameter Errors (1σ) for HERA.

constraints in the moderate scenario, in the optimistic scenario it has considerable discriminating power, similar to what can be achieved by jointly fitting all $z \leq 9$ data in the moderate model. This improvement in the effectiveness of the $z = 7$ measurement is due to an increased ability to access low and high k modes, which breaks degeneracies. With low and high k modes measurable, each redshift alone is already reasonably non-degenerate, and the main benefit (as far as degeneracy-breaking is concerned) in going to higher z is the opportunity to access new astrophysics with a slightly different set of degeneracies, rather than the opportunity to perform joint fits. Indeed, we see from the middle and bottom panels of Figure 8-16 that there are only minimal differences between the joint fit and the independent fits. In contrast, with the moderate model in Figure 8-15 we saw about a factor of four improvement in going from the latter to the former.

Figure 8-17 compares the ultimate performance of HERA for the three foreground scenarios, using all measured redshifts in a joint fit. (Note that our earlier emphasis on the differences between joint fits and independent fits was for pedagogical reasons only, since in practice there is no reason not to get the most out of one's data by performing a joint fit.) We see that even with the most pessimistic foreground model, our three parameters can be constrained to the 5% level. The ability to combine partially-coherent baselines in the moderate model results only in a modest improvement, but being able to work within the wedge in the optimistic case can suppress errors to the $\sim 1\%$ level. The final results are given in Table 8.3.

In closing, we see that a next-generation like HERA should be capable of delivering excellent constraints on astrophysical parameters during the EoR. These constraints will be particularly valuable, given that none of the parameters can be easily probed by other observations. However, a few qualifications are in order. First, the Fisher

matrix analysis performed here provides an accurate forecast of the errors only if the true parameter values are somewhat close to our fiducial ones. As an extreme example of how this could break down, suppose T_{vir} were actually 1000 K, as illustrated in the middle row of Figure 8-4. The result would be a high-redshift reionization scenario, one that would be difficult to probe to the precision demonstrated in this section, due to high thermal noise. Secondly, one's ability to extract interesting astrophysical quantities from a measurement of the power spectrum is only as good as one's ability to model the power spectrum. In this section, we assumed that `21cmFAST` is the "true" model of reionization. At the few-percent-level uncertainties given in Table 8.3, the measurement errors are better than or comparable to the scatter seen between different theoretical simulations [245]. Thus, there will likely need to be much feedback between theory and observation to make sense of a power spectrum measurement with HERA-level precision. Alternatively, given the small error bars seen here with a three-parameter model, it is likely that additional parameters can be added to one's power spectrum fits without sacrificing the ability to place constraints that are theoretically interesting. We leave the possibility of including additional parameters (many of which have smaller, subtler effects on the 21 cm power spectrum than the parameters examined here) for future work.

8.4 Conclusions

In order to explore the potential range of constraints that will come from the proposed next generation of 21 cm experiments (e.g. HERA and SKA), we used simple models for instruments, foregrounds, and reionization histories to encompass a broad range of possible scenarios. For an instrument model, we used the ~ 0.1 km^2 HERA concept array, and calculated power spectrum sensitivities using the method of [183]. To cover uncertainties in the foregrounds, we used three principal models. Both our pessimistic and moderate model assumes foregrounds occupy the wedge-like region in k-space observed by [182], extending 0.1 hMpc^{-1} past the analytic horizon limit. Thus, both cases are amenable to a strategy of foreground avoidance. What makes our

pessimistic model pessimistic is the decision to combine partially redundant baselines in an incoherent fashion, allowing one to completely sidestep the systematics highlighted by [89]. In the moderate model, these baselines are allowed to be combined coherently. Finally, in our optimistic model, the size of the wedge is reduced to a region defined by the FWHM of the primary beam. Given the small field of view of the dishes used in the HERA concept array, this model is effectively equivalent to one in which foreground removal techniques prove successful. Lastly, to cover the uncertainties in reionization history, we use 21cmFAST to generate power spectra for a wide range of uncertain parameters: the ionizing efficiency ζ, the minimum virial temperature of halos producing ionizing photos, T_{vir}, and the mean free path of ionizing photons through the IGM R_{mfp}.

Looking at predicted power spectrum measurements for these various scenarios yields the following conclusions:

- Even with no development of analysis techniques beyond those used in [173], an experiment with $\sim 0.1\,\mathrm{km}^2$ of collecting area can yield very high significance $\gtrsim 30\sigma$ detections of nearly any reionization power spectrum (cf. Figure 8-7).

- Developing techniques that allow for the coherent addition of partially redundant baselines can result in a small increase of additional power spectrum sensitivity. In this work, we find our moderate foreground removal model to increase sensitivities by $\sim 20\%$ over our most pessimistic scenario. Generally, we find that coherent combination of partially redundant baselines reduces thermal noise errors by $\sim 40\%$, so addressing this issue will be somewhat more important for smaller arrays that have not yet reached the sample variance dominated regime.

- With the sensitivities achievable with our moderate foreground model, the next generation of arrays will yield high significance detections of the EoR power spectra, and provide detailed characterization of the power spectrum shape over an order-of-magnitude in k ($k \sim 0.1$–$1.0h\mathrm{Mpc}^{-1}$). These sensitivity levels may even allow for direct imaging of the EoR on these scales.

- If successful, foreground removal algorithms can dramatically boost the sensitivity of 21 cm measurements. They are also the only way to open up the largest scales of the power spectrum, which can lead to new physical insight through observations of the generic "knee" feature.

- Although it will represent a major breakthrough for the 21 cm cosmology community, a low to moderate (~ 5–10σ) detection of the EoR power spectrum may not be able to conclusively identify the redshift of 50% ionization. One might expect otherwise, since the peak brightness of the power spectrum occurs near this ionization fraction. However, accounting for the steep rise in T_{sys} at low frequencies, shows that the rise-and-fall of the power spectrum versus redshift may not be conclusively measurable without a higher significance measurement, such as those possible with the HERA design.

- Going beyond power spectrum measurements to astrophysical parameter constraints, lower redshifts observations are particularly prone to parameter degeneracies. These can be partially broken by foreground removal from within the wedge region (allowing access to the lowest k modes). Alternatively, degeneracies can be broken by performing parameter fits over multiple redshifts simultaneously (which is equivalent to making use of information about the power spectrum's *evolution*). Higher redshifts ($z \geq 11$) are typically limited not by intrinsic degeneracies, but by high thermal noise (at least for a HERA-like array), and add relatively little to constraints on reionization.

- Assuming a fiducial 21cmFAST reionization model, a HERA-like array will be capable of constraining reionization parameters to $\sim 1\%$ uncertainty if foreground removal within the wedge proves possible, and to $\sim 5\%$ otherwise. The current generation of interferometers will struggle to provide precise constraints on reionization models; the sensitivity of a HERA-like array is necessary for this kind of science (for a quantitative comparison, see the appendix).

From this analysis, it is clear that for 21 cm studies to deliver the first conclusive scientific constraints on the Epoch of Reionization, arrays much larger than those

451

currently operational must be constructed. Advancements in analysis techniques to keep the EoR window free from contamination can contribute additional sensitivity, but the most dramatic gains on the analysis front will come from techniques that remove foreground emission and allow retrieval of modes from inside the wedge. This is not meant to disparage the wide range of foreground removal techniques already in the literature; rather, the impetus is on adapting these techniques for application to real data from the current and next generation of 21 cm experiments. The vast range of EoR science achievable under our optimistic, moderate, and even pessimistic foreground removal scenarios provides ample motivation for continuing these efforts.

8.A Appendix: Power Spectrum Sensitivities of Other 21 cm Experiments

In this appendix, we compare the power spectrum sensitivities and EoR parameter constraints of several 21 cm experiments. In particular, we consider the current generation experiments of PAPER [171], the MWA [220], and LOFAR [229], as well as a concept array for Phase 1 of the SKA based on the SKA System Baseline Design document (SKA-TEL-SKO-DD-001[12]). The instrument designs are summarized in Table 8.4, and the principal results are presented in Tables 8.5 and 8.6, which show the significance of the power spectrum measurements and constraints on EoR astrophysical parameters, respectively. Both calculations assume the fiducial EoR history shown in Figure 8-3. The significances in Table 8.5 assume only an 8 MHz band centered on the 50% ionization redshift of $z = 9.5$. The astrophysical constraints, however, assume information is collected over a wider band from $z = 7$–13; for instruments with smaller instantaneous bandwidths, the observing times will need to be adjusted accordingly.[13]

---

[12]http://www.skatelescope.org/wp-content/uploads/2012/07/SKA-TEL-SKO-DD-001-1_BaselineDesign1.pdf

[13]For the particulars of our fiducial EoR model, a significant fraction of the information comes from $z = 7$–9 (142–178 MHz) (see Figure 8-16), meaning that an experiment like the MWA with an instantaneous bandwidth of 30 MHz could nearly produce the results described here without a signficant correction for observing time. Of course, this assumes that the redshift of reionization is

Instrument	Number of Elements	Element Size (m²)	Collecting Area (m²)	Configuration
PAPER	132	9	1188	11×12 sparse grid
MWA	128	28	3584	Dense 100 m core with r^{-2} distribution beyond
LOFAR NL Core	48[a]	745	35,762	Dense 2 km core
HERA	547	154	84,238	Filled 200 m hexagon
SKA1 Low Core	866	962	833,190	Filled 270 m core with Gaussian distribution beyond

Table 8.4: Properties of Other 21 cm Experiments.

Instrument	Pessimistic	Moderate	Optimistic
PAPER	1.17	2.02	4.82
MWA	0.60	2.46	6.40
LOFAR NL Core	1.35	2.76	17.37
HERA	32.09	38.20	133.15
SKA1 Low Core	10.01	35.95	218.27

Table 8.5: Power spectrum measurement signifiance (number of σs) of other 21 cm experiments for each of the three foreground removal models.

	Pessimistic			Moderate			Optimistic		
Instrument	$\frac{\Delta T_{\mathrm{vir}}}{T_{\mathrm{vir,fid}}}$	$\frac{\Delta \zeta}{\zeta_{\mathrm{fid}}}$	$\frac{\Delta R_{\mathrm{mfp}}}{R_{\mathrm{mfp,fid}}}$	$\frac{\Delta T_{\mathrm{vir}}}{T_{\mathrm{vir,fid}}}$	$\frac{\Delta \zeta}{\zeta_{\mathrm{fid}}}$	$\frac{\Delta R_{\mathrm{mfp}}}{R_{\mathrm{mfp,fid}}}$	$\frac{\Delta T_{\mathrm{vir}}}{T_{\mathrm{vir,fid}}}$	$\frac{\Delta \zeta}{\zeta_{\mathrm{fid}}}$	$\frac{\Delta R_{\mathrm{mfp}}}{R_{\mathrm{mfp,fid}}}$
PAPER	1.444	1.168	1.507	1.260	1.013	1.294	0.272	0.179	0.140
MWA	4.419	3.479	4.555	0.757	0.568	0.731	0.231	0.152	0.119
LOFAR	1.538	1.251	1.515	0.719	0.565	0.675	0.069	0.046	0.039
HERA	0.072	0.051	0.047	0.062	0.044	0.039	0.011	0.007	0.005
SKA1	0.235	0.169	0.179	0.076	0.054	0.044	0.009	0.006	0.004

Table 8.6: Fractional errors on the reionization parameters achieveable with each instrument under the three foreground removal models, assuming all redshifts are analyzed jointly.

In order to compute the constraints achievable with other experiments, we apply the sensitivity calculation described in Section 8.2.1.1 to each of the five instruments under study. We note that this sensitivity calculation assumes a drift-scanning observing mode, with the limit of coherent sampling set by the size of the element primary beam. The MWA, LOFAR, and likely the SKA all have the capability of conducting a tracked scan to increase the coherent integration on a single patch of sky. Similarly, tracking can be used to move to declinations away from zenith if sample variance becomes the dominant source of error. A full study of the benefits of tracking versus draft scanning for power spectrum measurements is beyond the scope of this present work; rather, we assume all instruments operate in a drift-scanning mode for the clearest comparison with the fiducial results calculated for the HERA experiment. We therefore also assume that each telescope observes for the fiducial 6 hours per day for 180 days (1080 hours). Finally, we also assume that each array has a receiver temperature of 100 K. We discuss the important features of each instrument and the resultant constraints in turn; see the main text for a discussion of the HERA experiment.

1. **PAPER**: Our fiducial PAPER instrument is an 11×12 grid of PAPER dipoles modeled after the maximum redundancy arrays presented in [170]. In this configuration, the 3×3 m dipoles are spaced in 12 north-south columns separated by 16 m; within a column, the dipoles are spaced 4 m apart. In both our pessimistic and moderate scenarios, PAPER yields a non-detection of the fiducial 21 cm power spectrum. In the optimistic scenario, the array could yield a significant detection; however, the poor PSF of the maximum redundancy array is expected to present challenges to any foreground-removal strategy that would allow recovery from information inside the wedge [173]. Therefore, achieving the results of the optimistic scenario will be especially difficult for the PAPER experiment.

2. **MWA**: Our model MWA array uses the 128 antenna positions presented in

known a priori, and that the optimal band for constraints is actually the band observed.

[220]. Despite having nearly three times the collecting area of the PAPER array, we find the MWA yields a less significant detection in the pessimistic scenario. Poor sensitivity when partially redundant samples are combined incoherently is to be expected for the MWA. The pseudo-random configuration of the array produces essentially no instantaneously redundant samples, and so all redundancy comes from partial coherence. Therefore, one might expect the MWA to under-perform compared to the highly redundant PAPER array in this scenario. In the moderate and optimistic scenarios where partial redundancy yields sensitivity boosts the MWA outperforms the PAPER array.

3. **LOFAR**: To model the LOFAR array, we use the antenna positions presented in [229]. For the purposes of EoR power spectrum studies, we focus on the Netherlands core of the instrument, since baselines much longer than a few km contribute very little sensitivity. We also assume that LOFAR is operated in a mode where each sub-station of the HBA is correlated separately to increase the number of short baselines. However, the resultant sensitivities still show that LOFAR suffers from a lack of short baselines. Despite having a collecting area $\gtrsim$ 10 times larger than PAPER and the MWA, LOFAR still yields a non-detection of the EoR power spectrum in the pessimistic and moderate foreground removal scenarios. Only in the optimistic scenario where longer baselines contribute to the power spectrum measurements does LOFAR's collecting area result in a high-significance measurement. Preliminary results from the LOFAR experiment show significant progress in subtracting foregrounds to access modes inside the wedge [39, 244].

4. **SKA1-Low**: We model our SKA-Low Phase 1 instrument after the design parameters set out in the SKA System Baseline Design document, although the final design of the SKA is still subject to change. This document specifies that the array will consist of 911 35 m stations, with 866 stations in a core with a Gaussian distribution versus radius. This distribution is normalized to have 650 stations within a radius of 1 km. This density in fact yields a completely filled
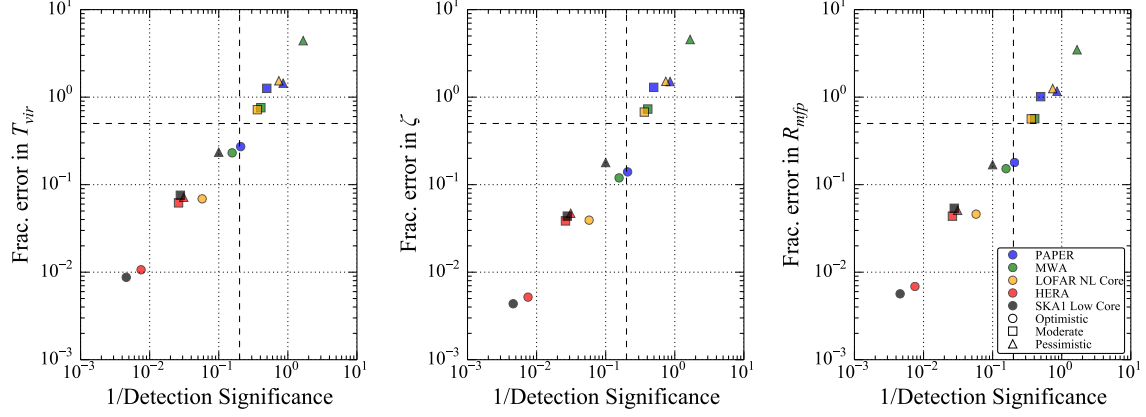
Figure 8-18: Fractional errors in astrophysical parameters shown as a function of (detection significance)$^{-1}$. Different instruments are shown in different colors (PAPER in blue; MWA in green; LOFAR in yellow; HERA in red; SKA in black), and different foreground scenarios are shown using different shapes (optimistic as circles; moderate as squares; pessimistic as triangles). The vertical dashed line delineates a 5σ detection of the power spectrum, while the horizontal dashed line delineates a parameter error of 50%. The tight correlations shown here suggest that the significance of a power spectrum detection can be used as a proxy for an instrument's ability to constrain astrophysical parameters.

aperture out to ~ 300 m, which we model as a close packed hexagon. This core gives the design some degree of instantaneous redundancy, a configuration that is still being explored for the final design of the instrument. We do not consider the 45 outriggers in our power spectrum sensitivity. Much like the case with PAPER and the MWA, the lower instantaneous redundancy of the SKA concept array results in a poorer performance than the highly redundant HERA array in the pessimistic scenario. However, in the moderate and optimistic scenarios this SKA concept design yields high sensitivity measurements, although not as high as might be expected from collecting area alone. This fact is once again due to the relatively small number of short spacings compared to the HERA array, resulting in similar performances for the two configurations in the moderate scenario. As with LOFAR, the SKA design shines in the optimistic scenario, producing a very high SNR measurement.

In all cases, we find that the fractional errors on the reionization parameters

(Table 8.6) scale very closely with the overall significance of the power spectrum measurement (Table 8.5). This is shown in Figure 8-18, where we plot the fractional errors on the reionization parameters against the reciprocal of the power spectrum detection significance. These two quantities are seen to be directly proportional to an excellent approximation, regardless of foreground scenario.[14] Therefore, while the power spectrum sensitivity of an array can be a strong function of an instrument's configuration, the resultant astrophysical constraints are fairly generalizable once the instrument sensitivity is known. This is strongly suggestive that the results in the main body of the paper can be easily extended to other instruments. Also noteworthy is the fact that current-generation instruments encroach on the lower-left regions (high detection significance; small parameter errors) of the plots in Figure 8-18 only for the optimistic foreground model. In contrast, the next-generation instruments (HERA and SKA) are clearly capable of delivering excellent scientific results even in the most pessimistic foreground scenario.

---

[14]We note that this is true only when all redshifts are analyzed jointly, where the errors are driven mainly by thermal noise. If the errors are instead dominated by parameter degeneracies (as is the case, for example, when only one redshift slice is measured), the tight linear correlation breaks down.

Chapter 9

Conclusion

This thesis focused on the development of and early results from the field of 21 cm cosmology, a new and potentially transformative probe of the universe during the first billion years after the big bang, a period of dramatic change we call the "Cosmic Dawn." Using the 21 cm transition of neutral hydrogen, large volumes of the early intergalactic medium can be mapped tomographically. This will open up an enormous and largely unexplored volume to precise tests of our astrophysical and cosmological models.

Realizing the promise of 21 cm cosmology will be extraordinarily difficult. We need very large radio telescopes observing for hundreds or thousands of hours just to achieve the necessary sensitivity to see the faint signal from neutral hydrogen. We also need to separate out the cosmological signal from astrophysical foregrounds, which are four or more orders of magnitude brighter, using our understanding of the physical processes that create them and the way they appear in our instruments.

I began this thesis in Chapter 1 explaining the current state of our understanding of the physics that underlies the Cosmic Dawn and the first stars, galaxies, and black holes that drove it. I then explained why the 21 cm signal can be detected and the physical processes that affect it. I also reviewed the observational challenges—especially that of bright foregrounds as seen through an interferometer—and surveyed the current and near future efforts to detect and characterize the cosmological signal.

The rest of the thesis was split into three parts. In Part I, *Novel Data Analysis*

Tools, I reproduced two theoretical papers that detailed analysis techniques designed for the rigorous and robust detection the power spectrum of 21 cm brightness temperature fluctuations. Chapter 2 focused on the acceleration of the previously published quadratic estimator method of Liu and Tegmark [120] to meet the challenges of large data volumes typically encountering in 21 cm power spectrum estimation. Chapter 3 relaxed one of the key assumptions in Chapter 2 to incorporate realistic chromatic effects on the point spread function—effects that create the characteristic "wedge" feature in the cylindrically averaged power spectrum. Like the previous chapter, Chapter 3 focused on showing how the statistical relationship between interferometric maps and the true sky can be modeled and propagated through the rest of the analysis in a computationally feasible way.

In Part II, *Early Results from New Telescopes*, I included three papers discussing scientific and technological progress toward eventually making a detection of the 21 cm signal. Chapters 4 and 5 presented upper limits on the 21 cm brightness temperature power spectrum using the Murchison Widefield Array, in its 32-tile prototype phase and then in its full 128-tile configuration. Both papers developed various methods for adapting the techniques from Part I to the challenges presented by real-world data, including calibration errors, incomplete Fourier sampling, radio frequency interference, and foreground residual modeling. Then, in Chapter 6, I reproduced a paper reporting on MITEoR, a technology demonstration array designed to test techniques—especially the redundant calibration of antenna gains and phases—for building highly scalable interferometers that may one day realize the full potential of 21 cm cosmology.

Finally in Part III, *The Cosmic Dawn on the Horizon*, I explored that potential in greater depth by reproducing two papers examining the scientific progress near future telescopes can make toward constraining the astrophysics behind the Cosmic Dawn. Chapter 7 looked at the possible effect of high redshift radio-loud active galactic nuclei on the 21 cm power spectrum. Depending on the population of those objects and the thermal history of the intergalactic medium, they could have a significant observable effect, especially at high redshift. Chapter 8 looked at the constraints the next

generation of 21 cm interferometers, especially the Hydrogen Epoch of Reionization Array (HERA), can place on the physics underlying reionization, by detecting and beginning to characterize the 21 cm power spectrum from the epoch of reionization.

Together this thesis represents six years of work toward the development of 21 cm cosmology as the next great cosmological probe—not to mention the tremendous contributions of collaborators and coauthors across three different telescope teams. And yet, to recall the words of Hubble, the thesis "ends on a note of uncertainty." We still haven't found the signal we're looking for. Even as we push to "the utmost limits of our telescopes" we find ourselves "measuring shadows" and pouring though vast quantities of data, beset by overwhelming foregrounds, and "searching among ghostly errors of measurement for landmarks that are scarcely more substantial." We've come a long way and we've still got a long way to go.

"The search will continue."

Bibliography

[1] T. Abel, G. L. Bryan, and M. L. Norman. The Formation of the First Star in the Universe. *Science*, 295:93–98, January 2002. doi: 10.1126/science.295.5552.93.

[2] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables.* Applied mathematics series. Dover Publications, 1964. ISBN 9780486612720. URL `http://books.google.com/books?id=hLqv8pHoOOAC`.

[3] P. A. R. Ade, R. W. Aikin, D. Barkats, S. J. Benton, C. A. Bischoff, J. J. Bock, J. A. Brevik, I. Buder, E. Bullock, C. D. Dowell, L. Duband, J. P. Filippini, S. Fliescher, S. R. Golwala, and M. Halpern. Detection of B-Mode Polarization at Degree Angular Scales by BICEP2. *Physical Review Letters*, 112(24):241101, June 2014. doi: 10.1103/PhysRevLett.112.241101.

[4] K. Ahn, A. Mesinger, M. A. Alvarez, and X. Chen. Probing First Galaxies and Their Impact on the Intergalactic Medium through the 21-cm Observation of the Cosmic Dawn with the SKA. *ArXiv e-prints*, January 2015.

[5] S. S. Ali, S. Bharadwaj, and B. Pandey. What will anisotropies in the clustering pattern in redshifted 21-cm maps tell us? *MNRAS*, 363:251–258, October 2005. doi: 10.1111/j.1365-2966.2005.09444.x.

[6] Z. S. Ali, A. R. Parsons, H. Zheng, J. C. Pober, A. Liu, J. E. Aguirre, R. F. Bradley, G. Bernardi, C. L. Carilli, C. Cheng, D. R. DeBoer, M. R. Dexter, J. Grobbelaar, J. Horrell, D. C. Jacobs, P. Klima, D. H. E. MacMahon, M. Maree, D. F. Moore, N. Razavi, I. I. Stefan, W. P. Walbrugh, and A. Walker. PAPER-64 Constraints on Reionization: The 21cm Power Spectrum at z=8.4. *ArXiv e-prints*, February 2015.

[7] D. C. Backer, J. Aguirre, J. D. Bowman, R. Bradley, C. L. Carilli, S. R. Furlanetto, L. J. Greenhill, J. N. Hewitt, C. Lonsdale, S. M. Ord, A. Parsons, and A. Whitney. Hera - hydrogen epoch of reionization arrays. In *US Astro2010 Decadal Survey "Cosmology and Fundamental Physics" Science Frontier Panel*, 2009.

[8] K. Bandura, G. E. Addison, M. Amiri, J. R. Bond, D. Campbell-Wilson, L. Connor, J.-F. Cliche, G. Davis, M. Deng, N. Denman, M. Dobbs, M. Fandino, K. Gibbs, A. Gilbert, M. Halpern, D. Hanna, A. D. Hincks, G. Hinshaw,

C. Höfer, P. Klages, T. L. Landecker, K. Masui, J. Mena Parra, L. B. Newburgh, U.-l. Pen, J. B. Peterson, A. Recnik, J. R. Shaw, K. Sigurdson, M. Sitwell, G. Smecher, R. Smegal, K. Vanderlinde, and D. Wiebe. Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9145 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 22, July 2014. doi: 10.1117/12.2054950.

[9] R. Barkana. Separating out the Alcock-Paczyński effect on 21-cm fluctuations. *MNRAS*, 372:259–264, October 2006. doi: 10.1111/j.1365-2966.2006.10882.x.

[10] R. Barkana. Studying the sources of cosmic reionization with 21-cm fluctuations. MNRAS, 397:1454–1463, August 2009. doi: 10.1111/j.1365-2966.2009.14929.x.

[11] R. Barkana and A. Loeb. In the beginning: the first sources of light and the reionization of the universe. Phys. Rep., 349:125–238, July 2001. doi: 10.1016/S0370-1573(01)00019-9.

[12] R. Barkana and A. Loeb. Detecting the Earliest Galaxies through Two New Sources of 21 Centimeter Fluctuations. *ApJ*, 626:1–11, June 2005. doi: 10.1086/429954.

[13] R. Barkana and A. Loeb. A Method for Separating the Physics from the Astrophysics of High-Redshift 21 Centimeter Fluctuations. *ApJ Lett.*, 624:L65–L68, May 2005. doi: 10.1086/430599.

[14] A. Beardsley and others. *in preparation.*, 2015.

[15] A. P. Beardsley, B. J. Hazelton, M. F. Morales, W. Arcus, D. Barnes, G. Bernardi, J. D. Bowman, F. H. Briggs, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. deSouza, D. Emrich, B. M. Gaensler, R. Goeke, L. J. Greenhill, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, E. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya Shankar, K. S. Srivani, J. Stevens, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. The EoR Sensitivity of the Murchison Widefield Array. *ArXiv e-prints*, April 2012.

[16] A. P. Beardsley, B. J. Hazelton, M. F. Morales, R. J. Capallo, R. Goeke, D. Emrich, C. J. Lonsdale, W. Arcus, D. Barnes, G. Bernardi, J. D. Bowman, J. D. Bunton, B. E. Corey, A. Deshpande, L. deSouza, B. M. Gaensler, L. J. Greenhill, D. Herne, J. N. Hewitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, E. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. U. Shankar, K. S.

464

Srivani, J. Stevens, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. A new layout optimization technique for interferometric arrays, applied to the Murchison Widefield Array. MNRAS, 425:1781–1788, September 2012. doi: 10.1111/j.1365-2966.2012.20878.x.

[17] R. H. Becker, X. Fan, R. L. White, M. A. Strauss, V. K. Narayanan, R. H. Lupton, J. E. Gunn, J. Annis, N. A. Bahcall, J. Brinkmann, A. J. Connolly, I. Csabai, P. C. Czarapata, M. Doi, T. M. Heckman, G. S. Hennessy, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, T. A. McKay, J. A. Munn, T. Nash, R. Nichol, J. R. Pier, G. T. Richards, D. P. Schneider, C. Stoughton, A. S. Szalay, A. R. Thakar, and D. G. York. Evidence for Reionization at z~6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. AJ, 122:2850–2857, December 2001. doi: 10.1086/324231.

[18] G. Bernardi, A. G. de Bruyn, M. A. Brentjens, B. Ciardi, G. Harker, V. Jelić, L. V. E. Koopmans, P. Labropoulos, A. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, S. Yatawatta, and S. Zaroubi. Foregrounds for observations of the cosmological 21 cm line. I. First Westerbork measurements of Galactic emission at 150 MHz in a low latitude field. A&A, 500:965–979, June 2009. doi: 10.1051/0004-6361/200911627.

[19] G. Bernardi, D. A. Mitchell, S. M. Ord, L. J. Greenhill, B. Pindor, R. B. Wayth, and J. S. B. Wyithe. Subtraction of point sources from interferometric radio images through an algebraic forward modelling scheme. MNRAS, 413:411–422, May 2011. doi: 10.1111/j.1365-2966.2010.18145.x.

[20] G. Bernardi, L. J. Greenhill, D. A. Mitchell, S. M. Ord, B. J. Hazelton, B. M. Gaensler, A. de Oliveira-Costa, M. F. Morales, N. Udaya Shankar, R. Subrahmanyan, R. B. Wayth, E. Lenc, C. L. Williams, W. Arcus, B. S. Arora, D. G. Barnes, J. D. Bowman, F. H. Briggs, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. deSouza, D. Emrich, R. Goeke, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, E. Morgan, D. Oberoi, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, K. S. Srivani, J. Stevens, S. J. Tingay, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, and J. S. B. Wyithe. A 189 MHz, 2400 deg^2 Polarization Survey with the Murchison Widefield Array 32-element Prototype. ApJ, 771:105, July 2013. doi: 10.1088/0004-637X/771/2/105.

[21] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson. Correcting direction-dependent gains in the deconvolution of radio interferometric images. A&A, 487:419–429, August 2008. doi: 10.1051/0004-6361:20079284.

[22] J. M. Bittner and A. Loeb. Measuring the redshift of reionization with a modest

array of low-frequency dipoles. J. Cosmology Astropart. Phys., 4:038, April 2011. doi: 10.1088/1475-7516/2011/04/038.

[23] J. R. Bond, A. H. Jaffe, and L. Knox. Estimating the power spectrum of the cosmic microwave background. Phys. Rev. D, 57:2117–2137, February 1998. doi: 10.1103/PhysRevD.57.2117.

[24] J. Bowman and A. E. E. Rogers. VHF-band RFI in Geographically Remote Areas. In *RFI Mitigation Workshop*, May 2010.

[25] J. D. Bowman, M. F. Morales, and J. N. Hewitt. The Sensitivity of First-Generation Epoch of Reionization Observatories and Their Potential for Differentiating Theoretical Power Spectra. *ApJ*, 638:20–26, February 2006. doi: 10.1086/498703.

[26] J. D. Bowman, M. F. Morales, and J. N. Hewitt. Constraints on Fundamental Cosmological Parameters with Upcoming Redshifted 21 cm Observations. *ApJ*, 661:1–9, May 2007. doi: 10.1086/516560.

[27] J. D. Bowman, A. E. E. Rogers, and J. N. Hewitt. Toward Empirical Constraints on the Global Redshifted 21 cm Brightness Temperature During the Epoch of Reionization. *ApJ*, 676:1–9, March 2008. doi: 10.1086/528675.

[28] J. D. Bowman, M. F. Morales, and J. N. Hewitt. Foreground Contamination in Interferometric Measurements of the Redshifted 21 cm Power Spectrum. *ApJ*, 695:183–199, April 2009. doi: 10.1088/0004-637X/695/1/183.

[29] J. D. Bowman, I. Cairns, D. L. Kaplan, T. Murphy, D. Oberoi, L. Staveley-Smith, W. Arcus, D. G. Barnes, G. Bernardi, F. H. Briggs, S. Brown, J. D. Bunton, A. J. Burgasser, R. J. Cappallo, S. Chatterjee, B. E. Corey, A. Coster, A. Deshpande, L. deSouza, D. Emrich, P. Erickson, R. F. Goeke, B. M. Gaensler, L. J. Greenhill, L. Harvey-Smith, B. J. Hazelton, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, L. D. Matthews, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. H. Morgan, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, T. Robishaw, A. E. E. Rogers, A. A. Roshi, J. E. Salah, R. J. Sault, N. U. Shankar, K. S. Srivani, J. B. Stevens, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. J. Williams, C. L. Williams, and J. S. B. Wyithe. Science with the Murchison Widefield Array. PASA, 30:e031, April 2013. doi: 10.1017/pas.2013.009.

[30] R. Braun. Understanding synthesis imaging dynamic range. A&A, 551:A91, March 2013. doi: 10.1051/0004-6361/201220257.

[31] V. Bromm, P. S. Coppi, and R. B. Larson. The Formation of the First Stars. I. The Primordial Star-forming Cloud. ApJ, 564:23–51, January 2002. doi: 10.1086/323947.

[32] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, and S. W. Randall. Detection of an Unidentified Emission Line in the Stacked X-Ray Spectrum of Galaxy Clusters. ApJ, 789:13, July 2014. doi: 10.1088/0004-637X/789/1/13.

[33] C. L. Carilli and S. Rawlings. Motivation, key science projects, standards and assumptions. New A Rev., 48:979–984, December 2004. doi: 10.1016/j.newar.2004.09.001.

[34] C. L. Carilli, N. Y. Gnedin, and F. Owen. H I 21 Centimeter Absorption beyond the Epoch of Reionization. ApJ, 577:22–30, September 2002. doi: 10.1086/342179.

[35] T. D. Carozzi and G. Woan. A generalized measurement equation and van Cittert-Zernike theorem for wide-field radio astronomical interferometry. MN-RAS, 395:1558–1568, May 2009. doi: 10.1111/j.1365-2966.2009.14642.x.

[36] P. Carroll and others. *in preparation.*, 2015.

[37] P. Carroll and others. *in preparation.*, 2015.

[38] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald. Baryon Acoustic Oscillation Intensity Mapping of Dark Energy. Phys. Rev. Lett.., 100(9): 091303–+, March 2008. doi: 10.1103/PhysRevLett.100.091303.

[39] E. Chapman, F. B. Abdalla, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans. Foreground removal using FASTICA: a showcase of LOFAR-EoR. MNRAS, 423:2518–2532, July 2012. doi: 10.1111/j.1365-2966.2012.21065.x.

[40] E. Chapman, F. B. Abdalla, J. Bobin, J.-L. Starck, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans. The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis. MNRAS, 429:165–176, February 2013. doi: 10.1093/mnras/sts333.

[41] J. Cho, A. Lazarian, and P. T. Timbie. A Technique for Foreground Subtraction in Redshifted 21 cm Observations. ApJ, 749:164, April 2012. doi: 10.1088/0004-637X/749/2/164.

[42] P. Christian and A. Loeb. Measuring the X-ray background in the reionization era with first generation 21 cm experiments. J. Cosmology Astropart. Phys., 9: 014, September 2013. doi: 10.1088/1475-7516/2013/09/014.

[43] B. Ciardi, P. Labropoulos, A. Maselli, R. Thomas, S. Zaroubi, L. Graziani, J. S. Bolton, G. Bernardi, M. Brentjens, A. G. de Bruyn, S. Daiboo, G. J. A. Harker, V. Jelic, S. Kazemi, L. V. E. Koopmans, O. Martinez, G. Mellema, A. R. Offringa, V. N. Pandey, J. Schaye, V. Veligatla, H. Vedantham, and

S. Yatawatta. Prospects for detecting the 21 cm forest from the diffuse intergalactic medium with LOFAR. MNRAS, 428:1755–1765, January 2013. doi: 10.1093/mnras/sts156.

[44] S. Clesse, L. Lopez-Honorez, C. Ringeval, H. Tashiro, and M. H. G. Tytgat. Background reionization history from omniscopes. Phys. Rev. D, 86(12):123506, December 2012. doi: 10.1103/PhysRevD.86.123506.

[45] A. S. Cohen, W. M. Lane, W. D. Cotton, N. E. Kassim, T. J. W. Lazio, R. A. Perley, J. J. Condon, and W. C. Erickson. The VLA Low-Frequency Sky Survey. AJ, 134:1245–1262, September 2007. doi: 10.1086/520719.

[46] Edward Collett. *Field guide to polarization*. SPIE Press, Bellingham, Wash, 2005. ISBN 0819458686.

[47] T. J. Cornwell, K. Golap, and S. Bhatnagar. The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm. *IEEE Journal of Selected Topics in Signal Processing*, 2:647–657, November 2008. doi: 10.1109/JSTSP.2008.2005290.

[48] W. D. Cotton. Lessons from the VLA Long Wavelength Sky Survey (VLSS). In N. Kassim, M. Perez, W. Junor, and P. Henning, editors, *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*, volume 345 of *Astronomical Society of the Pacific Conference Series*, page 337, December 2005.

[49] W. D. Cotton, J. J. Condon, R. A. Perley, N. Kassim, J. Lazio, A. Cohen, W. Lane, and W. C. Erickson. Beyond the isoplanatic patch in the VLA Low-frequency Sky Survey. In J. M. Oschmann, Jr., editor, *Ground-based Telescopes*, volume 5489 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 180–189, October 2004. doi: 10.1117/12.551298.

[50] A. Datta, J. D. Bowman, and C. L. Carilli. Bright Source Subtraction Requirements for Redshifted 21 cm Measurements. *ApJ*, 724:526–538, November 2010. doi: 10.1088/0004-637X/724/1/526.

[51] K. K. Datta, G. Mellema, Y. Mao, I. T. Iliev, P. R. Shapiro, and K. Ahn. Light-cone effect on the reionization 21-cm power spectrum. MNRAS, 424:1877–1891, August 2012. doi: 10.1111/j.1365-2966.2012.21293.x.

[52] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer. The Characterization of the Gamma-Ray Signal from the Central Milky Way: A Compelling Case for Annihilating Dark Matter. *ArXiv e-prints*, February 2014.

[53] A. de Oliveira-Costa, M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich. A model of diffuse Galactic radio emission from 10 MHz to 100 GHz. *MNRAS*, 388:247–260, July 2008. doi: 10.1111/j.1365-2966.2008.13376.x.

[54] A. de Oliveira-Costa, M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich. A model of diffuse Galactic radio emission from 10 MHz to 100 GHz. MNRAS, 388:247–260, July 2008. doi: 10.1111/j.1365-2966.2008.13376.x.

[55] G. de Zotti, M. Massardi, M. Negrello, and J. Wall. Radio and millimeter continuum surveys and their astrophysical implications. A&A Rev., 18:1–65, February 2010. doi: 10.1007/s00159-009-0026-0.

[56] C. DeGraf, T. Di Matteo, N. Khandai, R. Croft, J. Lopez, and V. Springel. Early black holes in cosmological simulations: luminosity functions and clustering behaviour. MNRAS, 424:1892–1898, August 2012. doi: 10.1111/j.1365-2966. 2012.21294.x.

[57] T. Di Matteo, R. Perna, T. Abel, and M. J. Rees. Radio Foregrounds for the 21 Centimeter Tomography of the Neutral Intergalactic Medium at High Redshifts. *ApJ*, 564:576–580, January 2002. doi: 10.1086/324293.

[58] J. S. Dillon, A. Liu, and M. Tegmark. A fast method for power spectrum and foreground analysis for 21 cm cosmology. Phys. Rev. D, 87(4):043005, February 2013. doi: 10.1103/PhysRevD.87.043005.

[59] J. S. Dillon, A. Liu, C. L. Williams, J. N. Hewitt, M. Tegmark, E. H. Morgan, A. M. Levine, M. F. Morales, S. J. Tingay, G. Bernardi, J. D. Bowman, F. H. Briggs, R. C. Cappallo, D. Emrich, D. A. Mitchell, D. Oberoi, T. Prabu, R. Wayth, and R. L. Webster. Overcoming real-world obstacles in 21 cm power spectrum estimation: A method demonstration and results from early Murchison Widefield Array data. Phys. Rev. D, 89(2):023002, January 2014. doi: 10.1103/PhysRevD.89.023002.

[60] J. S. Dillon, A. R. Neben, J. N. Hewitt, M. Tegmark, and others. *in preparation.*, 2015.

[61] J. S. Dillon, M. Tegmark, A. Liu, A. Ewall-Wice, J. N. Hewitt, M. F. Morales, A. R. Neben, A. R. Parsons, and H. Zheng. Mapmaking for precision 21 cm cosmology. Phys. Rev. D, 91(2):023002, January 2015. doi: 10.1103/PhysRevD. 91.023002.

[62] J. N. Douglas, F. N. Bash, F. A. Bozyan, G. W. Torrence, and C. Wolfe. The Texas Survey of Radio Sources Covering -35.5 degrees < declination < 71.5 degrees at 365 MHz. AJ, 111:1945, May 1996. doi: 10.1086/117932.

[63] A. Ewall-Wice, J. S. Dillon, A. Mesinger, and J. Hewitt. Detecting the 21 cm forest in the 21 cm power spectrum. MNRAS, 441:2476–2496, July 2014. doi: 10.1093/mnras/stu666.

[64] H. I. Ewen and E. M. Purcell. Observation of a line in the galactic radio spectrum: Radiation from galactic hydrogen at 1,420 mc./sec. *Nature*, 168 (4270):356, September 1951. doi: 10.1038/168356a0. URL `http://dx.doi. org/10.1038/168356a0`.

[65] X. Fan, M. A. Strauss, G. T. Richards, J. F. Hennawi, R. H. Becker, R. L. White, A. M. Diamond-Stanic, J. L. Donley, L. Jiang, J. S. Kim, M. Vestergaard, J. E. Young, J. E. Gunn, R. H. Lupton, G. R. Knapp, D. P. Schneider, W. N. Brandt, N. A. Bahcall, J. C. Barentine, J. Brinkmann, H. J. Brewington, M. Fukugita, M. Harvanek, S. J. Kleinman, J. Krzesinski, D. Long, E. H. Neilsen, Jr., A. Nitta, S. A. Snedden, and W. Voges. A Survey of z > 5.7 Quasars in the Sloan Digital Sky Survey. IV. Discovery of Seven Additional Quasars. AJ, 131:1203–1209, March 2006. doi: 10.1086/500296.

[66] G. B. Field. Excitation of the Hydrogen 21-CM Line. *Proceedings of the IRE*, 46:240–250, January 1958. doi: 10.1109/JRPROC.1958.286741.

[67] R. A. Fisher. *J. Roy. Stat. Soc.*, 98:39, 1935.

[68] G. Foster, J. Hickish, A. Magro, D. Price, and K. Zarb Adami. Implementation of a direct-imaging and FX correlator for the BEST-2 array. MNRAS, 439: 3180–3188, April 2014. doi: 10.1093/mnras/stu188.

[69] S. R. Furlanetto. The global 21-centimeter background from high redshifts. MNRAS, 371:867–878, September 2006. doi: 10.1111/j.1365-2966.2006.10725.x.

[70] S. R. Furlanetto and A. Loeb. The 21 Centimeter Forest: Radio Absorption Spectra as Probes of Minihalos before Reionization. ApJ, 579:1–9, November 2002. doi: 10.1086/342757.

[71] S. R. Furlanetto, A. Sokasian, and L. Hernquist. Observing the reionization epoch through 21-centimetre radiation. *MNRAS*, 347:187–195, January 2004. doi: 10.1111/j.1365-2966.2004.07187.x.

[72] S. R. Furlanetto, M. Zaldarriaga, and L. Hernquist. Statistical Probes of Reionization with 21 Centimeter Tomography. *ApJ*, 613:16–22, September 2004. doi: 10.1086/423028.

[73] S. R. Furlanetto, M. McQuinn, and L. Hernquist. Characteristic scales during reionization. MNRAS, 365:115–126, January 2006. doi: 10.1111/j.1365-2966. 2005.09687.x.

[74] S. R. Furlanetto, S. P. Oh, and F. H. Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. Phys. Rep., 433:181–301, October 2006. doi: 10.1016/j.physrep.2006.08.002.

[75] S. R. Furlanetto, A. Lidz, A. Loeb, M. McQuinn, J. R. Pritchard, P. R. Shapiro, M. A. Alvarez, D. C. Backer, J. D. Bowman, J. O. Burns, C. L. Carilli, R. Cen, A. Cooray, N. Gnedin, L. J. Greenhill, Z. Haiman, J. N. Hewitt, and et al. Cosmology from the Highly-Redshifted 21 cm Line. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, pages 82–+, 2009.

[76] M. A. Garrett. Building LOFAR - status update. *ArXiv e-prints*, September 2009.

[77] A. Ghosh, J. Prasad, S. Bharadwaj, S. S. Ali, and J. N. Chengalur. Characterizing foreground for redshifted 21 cm radiation: 150 MHz Giant Metrewave Radio Telescope observations. MNRAS, 426:3295–3314, November 2012. doi: 10.1111/j.1365-2966.2012.21889.x.

[78] L. Gleser, A. Nusser, and A. J. Benson. Decontamination of cosmological 21-cm maps. *MNRAS*, 391:383–398, November 2008. doi: 10.1111/j.1365-2966.2008. 13897.x.

[79] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, 622: 759–771, April 2005. doi: 10.1086/427976.

[80] R. M. Gray. Toeplitz and Circulant Matrices: A Review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2006.

[81] L. J. Greenhill and G. Bernardi. HI Epoch of Reionization Arrays. *ArXiv e-prints*, January 2012.

[82] B. Greig and A. Mesinger. 21CMMC: An MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21cm signal. *ArXiv e-prints*, January 2015.

[83] M. R. Griffith, A. E. Wright, B. F. Burke, and R. D. Ekers. The Parkes-MIT-NRAO (PMN) surveys. 6: Source catalog for the equatorial survey (-9.5 deg less than delta less than +10.0 deg). ApJS, 97:347–453, April 1995. doi: 10.1086/192146.

[84] Z. Haiman, A. A. Thoul, and A. Loeb. Cosmological Formation of Low-Mass Objects. ApJ, 464:523, June 1996. doi: 10.1086/177343.

[85] Z. Haiman, E. Quataert, and G. C. Bower. Modeling the Counts of Faint Radio-Loud Quasars: Constraints on the Supermassive Black Hole Population and Predictions for High Redshift. ApJ, 612:698–705, September 2004. doi: 10.1086/422834.

[86] G. Harker, S. Zaroubi, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, V. Jelić, L. V. E. Koopmans, P. Labropoulos, G. Mellema, A. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, and S. Yatawatta. Non-parametric foreground subtraction for 21-cm epoch of reionization experiments. MNRAS, 397:1138–1152, August 2009. doi: 10.1111/j.1365-2966.2009.15081.x.

[87] G. Harker, S. Zaroubi, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, V. Jelić, L. V. E. Koopmans, P. Labropoulos, G. Mellema, A. Offringa, V. N.

Pandey, A. H. Pawlik, J. Schaye, R. M. Thomas, and S. Yatawatta. Power spectrum extraction for redshifted 21-cm Epoch of Reionization experiments: the LOFAR case. MNRAS, 405:2492–2504, July 2010. doi: 10.1111/j.1365-2966. 2010.16628.x.

[88] C. G. T. Haslam, C. J. Salter, H. Stoffel, and W. E. Wilson. A 408 MHz all-sky continuum survey. II - The atlas of contour maps. *A&AS*, 47:1, January 1982.

[89] B. J. Hazelton, M. F. Morales, and I. S. Sullivan. The Fundamental Multi-baseline Mode-mixing Foreground in 21 cm Epoch of Reionization Observations. ApJ, 770:156, June 2013. doi: 10.1088/0004-637X/770/2/156.

[90] M. R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49: 409–436, 1952.

[91] G. Hinshaw, D. N. Spergel, L. Verde, R. S. Hill, S. S. Meyer, C. Barnes, C. L. Bennett, M. Halpern, N. Jarosik, A. Kogut, E. Komatsu, M. Limon, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: The Angular Power Spectrum. ApJS, 148:135–159, September 2003. doi: 10.1086/377225.

[92] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS, 208:19, October 2013. doi: 10.1088/0067-0049/208/2/19.

[93] C. M. Hirata. Wouthuysen-Field coupling strength and application to high-redshift 21-cm radiation. MNRAS, 367:259–274, March 2006. doi: 10.1111/j. 1365-2966.2005.09949.x.

[94] J. A. Högbom. Aperture Synthesis with a Non-Regular Distribution of Inter-ferometer Baselines. A&AS, 15:417, June 1974.

[95] D. W. Hogg. Distance measures in cosmology. *ArXiv Astrophysics e-prints*, May 1999.

[96] J. Horgan. *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*. Helix Books. Addison-Wesley, 1996.

[97] N. Hurley-Walker, J. Morgan, R. B. Wayth, M. Bell, G. Bernardi, R. Bhat, F. Frank Briggs, Deshpande A. A., A. Ewall-Wice, L. Feng, B. Hazelton, L. Hindson, D. Jacobs, D. Kaplan, N. Kudryavtseva, E. Lenc, B. McKinley, D. Mitchell, B. Pindor, P. Procopio, D. Oberoi, M. Offringa, S. Ord, J. Riding, J. Bowman, R. Cappallo, B. Corey, D. Emrich, B. Gaensler, R. Goeke, L. Green-hill, J. Hewitt, M. Johnston-Hollitt, J. Kasper, E. Kratzenberg, C. Lonsdale,

M. Lynch, R. McWhirter, M. Morales, E. Morgan, T. Prabu, A. Rogers, A. Roshi, U. Shankar, K. Srivani, R. Subrahmanyan, S. Tingay, M. Waterson, R. Webster, A. Whitney, A. Williams, and C. Williams. The Murchison Widefield Array Commissioning Survey: A Low-Frequency Catalogue of 14 110 Compact Radio Sources over 6 100 Square Degrees. PASA, 31:e045, November 2014. doi: 10.1017/pasa.2014.40.

[98] I. T. Iliev, P. R. Shapiro, A. Ferrara, and H. Martel. On the Direct Detectability of the Cosmic Dark Ages: 21 Centimeter Emission from Minihalos. *ApJL*, 572: L123–L126, June 2002. doi: 10.1086/341869.

[99] I. T. Iliev, G. Mellema, P. R. Shapiro, U.-L. Pen, Y. Mao, J. Koda, and K. Ahn. Can 21-cm observations discriminate between high-mass and low-mass galaxies as reionization sources? MNRAS, 423:2222–2253, July 2012. doi: 10.1111/j. 1365-2966.2012.21032.x.

[100] H. T. Intema, S. van der Tol, W. D. Cotton, A. S. Cohen, I. M. van Bemmel, and H. J. A. Röttgering. Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. A&A, 501:1185–1205, July 2009. doi: 10.1051/0004-6361/200811094.

[101] Ž. Ivezić, K. Menou, G. R. Knapp, M. A. Strauss, R. H. Lupton, D. E. Vanden Berk, G. T. Richards, C. Tremonti, M. A. Weinstein, S. Anderson, N. A. Bahcall, R. H. Becker, M. Bernardi, M. Blanton, D. Eisenstein, X. Fan, D. Finkbeiner, K. Finlator, J. Frieman, J. E. Gunn, P. B. Hall, R. S. J. Kim, A. Kinkhabwala, V. K. Narayanan, C. M. Rockosi, D. Schlegel, D. P. Schneider, I. Strateva, M. SubbaRao, A. R. Thakar, W. Voges, R. L. White, B. Yanny, J. Brinkmann, M. Doi, M. Fukugita, G. S. Hennessy, J. A. Munn, R. C. Nichol, and D. G. York. Optical and Radio Properties of Extragalactic Sources Observed by the FIRST Survey and the Sloan Digital Sky Survey. AJ, 124:2364–2400, November 2002. doi: 10.1086/344069.

[102] D. Jacobs and others. The Murchison Widefield Array Epoch of Reionization Pipelines. *in preparation.*, 2015.

[103] D. C. Jacobs, J. Bowman, and J. E. Aguirre. The Precision and Accuracy of Early Epoch of Reionization Foreground Models: Comparing MWA and PAPER 32-antenna Source Catalogs. ApJ, 769:5, May 2013. doi: 10.1088/ 0004-637X/769/1/5.

[104] D. C. Jacobs, A. R. Parsons, J. E. Aguirre, Z. Ali, J. Bowman, R. F. Bradley, C. L. Carilli, D. R. DeBoer, M. R. Dexter, N. E. Gugliucci, P. Klima, D. H. E. MacMahon, J. R. Manley, D. F. Moore, J. C. Pober, I. I. Stefan, and W. P. Walbrugh. A Flux Scale for Southern Hemisphere 21 cm Epoch of Reionization Experiments. ApJ, 776:108, October 2013. doi: 10.1088/0004-637X/776/2/108.

[105] D. C. Jacobs, J. C. Pober, A. R. Parsons, J. E. Aguirre, Z. Ali, J. Bowman, R. F. Bradley, C. L. Carilli, D. R. DeBoer, M. R. Dexter, N. E. Gugliucci,

P. Klima, A. Liu, D. H. E. MacMahon, J. R. Manley, D. F. Moore, I. I. Stefan, and W. P. Walbrugh. Multi-redshift limits on the 21cm power spectrum from PAPER. *ArXiv e-prints*, August 2014.

[106] N. Jarosik, C. Barnes, M. R. Greason, R. S. Hill, M. R. Nolta, N. Odegard, J. L. Weiland, R. Bean, C. L. Bennett, O. Doré, M. Halpern, G. Hinshaw, A. Kogut, E. Komatsu, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Beam Profiles, Data Processing, Radiometer Characterization, and Systematic Error Limits. ApJS, 170:263–287, June 2007. doi: 10.1086/513697.

[107] V. Jelić, S. Zaroubi, P. Labropoulos, R. M. Thomas, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, G. Harker, L. V. E. Koopmans, V. N. Pandey, J. Schaye, and S. Yatawatta. Foreground simulations for the LOFAR-epoch of reionization experiment. *MNRAS*, 389:1319–1335, September 2008. doi: 10.1111/j.1365-2966.2008.13634.x.

[108] V. Jelić, S. Zaroubi, P. Labropoulos, G. Bernardi, A. G. de Bruyn, and L. V. E. Koopmans. Realistic simulations of the Galactic polarized foreground: consequences for 21-cm reionization detection experiments. MNRAS, 409:1647–1659, December 2010. doi: 10.1111/j.1365-2966.2010.17407.x.

[109] V. Jelić, A. G. de Bruyn, M. Mevius, F. B. Abdalla, K. M. B. Asad, G. Bernardi, M. A. Brentjens, S. Bus, E. Chapman, B. Ciardi, S. Daiboo, E. R. Fernandez, A. Ghosh, G. Harker, H. Jensen, S. Kazemi, L. V. E. Koopmans, P. Labropoulos, O. Martinez-Rubi, G. Mellema, A. R. Offringa, V. N. Pandey, A. H. Patil, R. M. Thomas, H. K. Vedantham, V. Veligatla, S. Yatawatta, S. Zaroubi, A. Alexov, J. Anderson, I. M. Avruch, R. Beck, M. E. Bell, M. J. Bentum, P. Best, A. Bonafede, J. Bregman, F. Breitling, J. Broderick, W. N. Brouw, M. Brüggen, H. R. Butcher, J. E. Conway, F. de Gasperin, E. de Geus, A. Deller, R.-J. Dettmar, S. Duscha, J. Eislöffel, D. Engels, H. Falcke, R. A. Fallows, R. Fender, C. Ferrari, W. Frieswijk, M. A. Garrett, J. Grießmeier, A. W. Gunst, J. P. Hamaker, T. E. Hassall, M. Haverkorn, G. Heald, J. W. T. Hessels, M. Hoeft, J. Hörandel, A. Horneffer, A. van der Horst, M. Iacobelli, E. Juette, A. Karastergiou, V. I. Kondratiev, M. Kramer, M. Kuniyoshi, G. Kuper, J. van Leeuwen, P. Maat, G. Mann, D. McKay-Bukowski, J. P. McKean, H. Munk, A. Nelles, M. J. Norden, H. Paas, M. Pandey-Pommier, G. Pietka, R. Pizzo, A. G. Polatidis, W. Reich, H. Röttgering, A. Rowlinson, A. M. M. Scaife, D. Schwarz, M. Serylak, O. Smirnov, M. Steinmetz, A. Stewart, M. Tagger, Y. Tang, C. Tasse, S. ter Veen, S. Thoudam, C. Toribio, R. Vermeulen, C. Vocks, R. J. van Weeren, R. A. M. J. Wijers, S. J. Wijnholds, O. Wucknitz, and P. Zarka. Initial LOFAR observations of epoch of reionization windows. II. Diffuse polarized emission in the ELAIS-N1 field. A&A, 568:A101, August 2014. doi: 10.1051/0004-6361/201423998.

[110] V. Jelic, B. Ciardi, E. Fernandez, H. Tashiro, and D. Vrbanec. SKA - EoR correlations and cross-correlations: kSZ, radio galaxies, and NIR background. *ArXiv e-prints*, January 2015.

[111] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18, February 2011. doi: 10.1088/0067-0049/192/2/18.

[112] Thomas Samuel Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1962 1962.

[113] P. La Plante, N. Battaglia, A. Natarajan, J. B. Peterson, H. Trac, R. Cen, and A. Loeb. Reionization on Large Scales. IV. Predictions for the 21 cm Signal Incorporating the Light Cone Effect. ApJ, 789:31, July 2014. doi: 10.1088/0004-637X/789/1/31.

[114] M. I. Large, B. Y. Mills, A. G. Little, D. F. Crawford, and J. M. Sutton. The Molonglo Reference Catalogue of Radio Sources. MNRAS, 194:693–+, February 1981.

[115] D. Lee. Fast multiplication of a recursive block Toeplitz matrix by a vector and its application. *Journal of Complexity*, 2(4):295–305, December 1986. doi: 10.1016/0885-064X(86)90007-5.

[116] F. Li, T. J. Cornwell, and F. de Hoog. The application of compressive sampling to radio astronomy. I. Deconvolution. A&A, 528:A31, April 2011. doi: 10.1051/0004-6361/201015045.

[117] A. Lidz, O. Zahn, M. McQuinn, M. Zaldarriaga, and L. Hernquist. Detecting the Rise and Fall of 21 cm Fluctuations with the Murchison Widefield Array. ApJ, 680:962–974, June 2008. doi: 10.1086/587618.

[118] A. Lidz, O. Zahn, S. R. Furlanetto, M. McQuinn, L. Hernquist, and M. Zaldarriaga. Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum. ApJ, 690:252–266, January 2009. doi: 10.1088/0004-637X/690/1/252.

[119] A. Liu. *From theoretical promise to observational reality: calibration, foreground subtraction, and signal extraction in hydrogen cosmology*. PhD thesis, Massachusetts Institute of Technology, June 2012.

[120] A. Liu and M. Tegmark. A method for 21 cm power spectrum estimation in the presence of foregrounds. Phys. Rev. D, 83(10):103006, May 2011. doi: 10.1103/PhysRevD.83.103006.

[121] A. Liu and M. Tegmark. How well can we measure and understand foregrounds with 21-cm experiments? MNRAS, 419:3491–3504, February 2012. doi: 10.1111/j.1365-2966.2011.19989.x.

[122] A. Liu, M. Tegmark, J. Bowman, J. Hewitt, and M. Zaldarriaga. An improved method for 21-cm foreground removal. *MNRAS*, 398:401–406, September 2009. doi: 10.1111/j.1365-2966.2009.15156.x.

[123] A. Liu, M. Tegmark, and M. Zaldarriaga. Will point sources spoil 21-cm tomography? *MNRAS*, 394:1575–1587, April 2009. doi: 10.1111/j.1365-2966.2009.14426.x.

[124] A. Liu, M. Tegmark, S. Morrison, A. Lutomirski, and M. Zaldarriaga. Precision calibration of radio interferometers using redundant baselines. MNRAS, 408: 1029–1050, October 2010. doi: 10.1111/j.1365-2966.2010.17174.x.

[125] A. Liu, A. R. Parsons, and C. M. Trott. Epoch of reionization window. I. Mathematical formalism. Phys. Rev. D, 90(2):023018, July 2014. doi: 10.1103/PhysRevD.90.023018.

[126] A. Liu, A. R. Parsons, and C. M. Trott. Epoch of reionization window. II. Statistical methods for foreground wedge reduction. Phys. Rev. D, 90(2):023019, July 2014. doi: 10.1103/PhysRevD.90.023019.

[127] A. Loeb and S. R. Furlanetto. *The First Galaxies In The Universe*. Princeton University Press, Princeton, NJ, 2013.

[128] A. Loeb and M. Zaldarriaga. Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations through 21cm Tomography Prior to the Epoch of Structure Formation. Phys. Rev. Lett., 92(21):211301–+, May 2004. doi: 10.1103/PhysRevLett.92.211301.

[129] C. J. Lonsdale, R. J. Cappallo, M. F. Morales, F. H. Briggs, L. Benkevitch, J. D. Bowman, J. D. Bunton, S. Burns, B. E. Corey, L. Desouza, S. S. Doeleman, M. Derome, A. Deshpande, M. R. Gopala, L. J. Greenhill, D. E. Herne, J. N. Hewitt, and et al. The Murchison Widefield Array: Design Overview. *IEEE Proceedings*, 97:1497–1506, August 2009. doi: 10.1109/JPROC.2009.2017564.

[130] R. Maartens, F. B. Abdalla, M. Jarvis, M. G. Santos, and f. t. SKA Cosmology SWG. Cosmology with the SKA – overview. *ArXiv e-prints*, January 2015.

[131] K. J. Mack and D. H. Wesley. Primordial black holes in the Dark Ages: Observational prospects for future 21cm surveys. *ArXiv e-prints*, May 2008.

[132] K. J. Mack and J. S. B. Wyithe. Detecting the redshifted 21 cm forest during reionization. MNRAS, 425:2988–3001, October 2012. doi: 10.1111/j.1365-2966.2012.21561.x.

[133] P. Madau and M. Dickinson. Cosmic Star-Formation History. ARA&A, 52: 415–486, August 2014. doi: 10.1146/annurev-astro-081811-125615.

[134] P. Madau, A. Meiksin, and M. J. Rees. 21 Centimeter Tomography of the Intergalactic Medium at High Redshift. *ApJ*, 475:429–+, February 1997. doi: 10.1086/303549.

[135] Y. Mao, M. Tegmark, M. McQuinn, M. Zaldarriaga, and O. Zahn. How accurately can 21cm tomography constrain cosmology? Phys. Rev. D, 78(2): 023529–+, July 2008. doi: 10.1103/PhysRevD.78.023529.

[136] V. R. Marthi and J. Chengalur. Non-linear redundancy calibration. MNRAS, 437:524–531, January 2014. doi: 10.1093/mnras/stt1902.

[137] K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek. Measurement of 21 cm Brightness Fluctuations at z ~ 0.8 in Cross-correlation. ApJ, 763:L20, January 2013. doi: 10.1088/2041-8205/763/1/L20.

[138] M. McQuinn. Constraints on X-ray emissions from the reionization era. MNRAS, 426:1349–1360, October 2012. doi: 10.1111/j.1365-2966.2012.21792.x.

[139] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto. Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization. *ApJ*, 653:815–834, December 2006. doi: 10.1086/505167.

[140] M. McQuinn, L. Hernquist, M. Zaldarriaga, and S. Dutta. Studying reionization with Lyα emitters. MNRAS, 381:75–96, October 2007. doi: 10.1111/j. 1365-2966.2007.12085.x.

[141] M. McQuinn, A. Lidz, O. Zahn, S. Dutta, L. Hernquist, and M. Zaldarriaga. The morphology of HII regions during reionization. MNRAS, 377:1043–1063, May 2007. doi: 10.1111/j.1365-2966.2007.11489.x.

[142] A. Mesinger and M. Dijkstra. Ultraviolet radiative feedback during the advanced stages of reionization. MNRAS, 390:1071–1080, November 2008. doi: 10.1111/j.1365-2966.2008.13776.x.

[143] A. Mesinger and S. Furlanetto. Efficient Simulations of Early Structure Formation and Reionization. ApJ, 669:663–675, November 2007. doi: 10.1086/521806.

[144] A. Mesinger, S. Furlanetto, and R. Cen. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. MNRAS, 411:955–972, February 2011. doi: 10.1111/j.1365-2966.2010.17731.x.

[145] A. Mesinger, M. McQuinn, and D. N. Spergel. The kinetic Sunyaev-Zel'dovich signal from inhomogeneous reionization: a parameter space study. MNRAS, 422:1403–1417, May 2012. doi: 10.1111/j.1365-2966.2012.20713.x.

[146] A. Mesinger, A. Ferrara, and D. S. Spiegel. Signatures of X-rays in the early Universe. MNRAS, 431:621–637, May 2013. doi: 10.1093/mnras/stt198.

[147] A. Mesinger, A. Ewall-Wice, and J. Hewitt. Reionization and beyond: detecting the peaks of the cosmological 21 cm signal. MNRAS, 439:3262–3274, April 2014. doi: 10.1093/mnras/stu125.

[148] S. Mineo, M. Gilfanov, B. D. Lehmer, G. E. Morrison, and R. Sunyaev. X-ray emission from star-forming galaxies - III. Calibration of the L_X-SFR relation up to redshift z $\approx$ 1.3. MNRAS, 437:1698–1707, January 2014. doi: 10.1093/mnras/stt1999.

[149] D. A. Mitchell, L. J. Greenhill, R. B. Wayth, R. J. Sault, C. J. Lonsdale, R. J. Cappallo, M. F. Morales, and S. M. Ord. Real-Time Calibration of the Murchison Widefield Array. *IEEE Journal of Selected Topics in Signal Processing*, 2: 707–717, November 2008. doi: 10.1109/JSTSP.2008.2005327.

[150] D. F. Moore, J. E. Aguirre, A. R. Parsons, D. C. Jacobs, and J. C. Pober. The Effects of Polarized Foregrounds on 21 cm Epoch of Reionization Power Spectrum Measurements. ApJ, 769:154, June 2013. doi: 10.1088/0004-637X/769/2/154.

[151] M. F. Morales. Power Spectrum Sensitivity and the Design of Epoch of Reionization Observatories. *ApJ*, 619:678–683, February 2005. doi: 10.1086/426730.

[152] M. F. Morales. Enabling Next-Generation Dark Energy and Epoch of Reionization Radio Observatories with the MOFF Correlator. PASP, 123:1265–1272, November 2011. doi: 10.1086/663092.

[153] M. F. Morales and J. Hewitt. Toward Epoch of Reionization Measurements with Wide-Field Radio Observations. *ApJ*, 615:7–18, November 2004. doi: 10.1086/424437.

[154] M. F. Morales and J. S. B. Wyithe. Reionization and Cosmology with 21-cm Fluctuations. ARA&A, 48:127–171, September 2010. doi: 10.1146/annurev-astro-081309-130936.

[155] M. F. Morales, J. D. Bowman, and J. N. Hewitt. Improving Foreground Subtraction in Statistical Observations of 21 cm Emission from the Epoch of Reionization. *ApJ*, 648:767–773, September 2006. doi: 10.1086/506135.

[156] M. F. Morales, B. Hazelton, I. Sullivan, and A. Beardsley. Four Fundamental Foreground Power Spectrum Shapes for 21 cm Cosmology Observations. ApJ, 752:137, June 2012. doi: 10.1088/0004-637X/752/2/137.

[157] A. R. Neben, R. F. Bradley, J. N. Hewitt, G. Bernardi, J. D. Bowman, F. Briggs, R. J. Cappallo, A. A. Deshpande, R. Goeke, L. J. Greenhill, B. J. Hazelton, M. Johnston-Hollitt, D. L. Kaplan, C. J. Lonsdale, S. R. McWhirter, D. A.

Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, N. Udaya Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, R. L. Webster, A. Williams, and C. L. Williams. Measuring Phased-Array Antenna Beampatterns with High Dynamic Range for the Murchison Widefield Array using 137 MHz ORBCOMM Satellites. *ArXiv e-prints*, May 2015.

[158] L. B. Newburgh, G. E. Addison, M. Amiri, K. Bandura, J. R. Bond, L. Connor, J.-F. Cliche, G. Davis, M. Deng, N. Denman, M. Dobbs, M. Fandino, H. Fong, K. Gibbs, A. Gilbert, E. Griffin, M. Halpern, D. Hanna, A. D. Hincks, G. Hinshaw, C. Höfer, P. Klages, T. Landecker, K. Masui, J. M. Parra, U.-L. Pen, J. Peterson, A. Recnik, J. R. Shaw, K. Sigurdson, M. Sitwell, G. Smecher, R. Smegal, K. Vanderlinde, and D. Wiebe. Calibrating CHIME: a new radio interferometer to probe dark energy. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9145 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 4, July 2014. doi: 10.1117/12.2056962.

[159] J. E. Noordam. LOFAR calibration challenges. In J. M. Oschmann Jr., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5489 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 817–825, October 2004. doi: 10.1117/12.544262.

[160] P. Noorishad, S. J. Wijnholds, A. van Ardenne, and J. M. van der Hulst. Redundancy calibration of phased-array stations. A&A, 545:A108, September 2012. doi: 10.1051/0004-6361/201219087.

[161] A. Nusser. The Alcock-Paczyński test in redshifted 21-cm maps. *MNRAS*, 364: 743–750, December 2005. doi: 10.1111/j.1365-2966.2005.09603.x.

[162] A. R. Offringa, B. McKinley, N. Hurley-Walker, F. H. Briggs, R. B. Wayth, D. L. Kaplan, M. E. Bell, L. Feng, A. R. Neben, J. D. Hughes, J. Rhee, T. Murphy, N. D. R. Bhat, G. Bernardi, J. D. Bowman, R. J. Cappallo, B. E. Corey, A. A. Deshpande, D. Emrich, A. Ewall-Wice, B. M. Gaensler, R. Goeke, L. J. Greenhill, B. J. Hazelton, L. Hindson, M. Johnston-Hollitt, D. C. Jacobs, J. C. Kasper, E. Kratzenberg, E. Lenc, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, N. Kudryavtseva, D. Oberoi, S. M. Ord, B. Pindor, P. Procopio, T. Prabu, J. Riding, D. A. Roshi, N. U. Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, and C. L. Williams. WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy. MNRAS, 444: 606–619, October 2014. doi: 10.1093/mnras/stu1368.

[163] A. R. Offringa, R. B. Wayth, N. Hurley-Walker, D. L. Kaplan, N. Barry, A. P. Beardsley, M. E. Bell, G. Bernardi, J. D. Bowman, F. Briggs, J. R. Callingham, R. J. Cappallo, P. Carroll, A. A. Deshpande, J. S. Dillon, K. S. Dwarakanath, A. Ewall-Wice, L. Feng, B.-Q. For, B. M. Gaensler, L. J. Greenhill, P. Hancock, B. J. Hazelton, J. N. Hewitt, L. Hindson, D. C. Jacobs, M. Johnston-Hollitt,

A. D. Kapińska, H.-S. Kim, P. Kittiwisit, E. Lenc, J. Line, A. Loeb, C. J. Lonsdale, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, J. Morgan, A. R. Neben, D. Oberoi, S. M. Ord, S. Paul, B. Pindor, J. C. Pober, T. Prabu, P. Procopio, J. Riding, N. Udaya Shankar, S. Sethi, K. S. Srivani, L. Staveley-Smith, R. Subrahmanyan, I. S. Sullivan, M. Tegmark, N. Thyagarajan, S. J. Tingay, C. M. Trott, R. L. Webster, A. Williams, C. L. Williams, C. Wu, J. S. Wyithe, and Q. Zheng. The low-frequency environment of the Murchison Widefield Array: radio-frequency interference analysis and mitigation. *ArXiv e-prints*, January 2015.

[164] S. P. Oh, D. N. Spergel, and G. Hinshaw. An Efficient Technique to Determine the Power Spectrum from Cosmic Microwave Background Sky Maps. ApJ, 510: 551–563, January 1999. doi: 10.1086/306629.

[165] T. Okamoto, L. Gao, and T. Theuns. Mass loss of galaxies due to an ultraviolet background. MNRAS, 390:920–928, November 2008. doi: 10.1111/j.1365-2966. 2008.13830.x.

[166] G. Paciga, T.-C. Chang, Y. Gupta, R. Nityanada, J. Odegova, U.-L. Pen, J. B. Peterson, J. Roy, and K. Sigurdson. The GMRT Epoch of Reionization experiment: a new upper limit on the neutral hydrogen power spectrum at $z \approx 8.6$. MNRAS, 413:1174–1183, May 2011. doi: 10.1111/j.1365-2966.2011.18208.x.

[167] G. Paciga, J. G. Albert, K. Bandura, T.-C. Chang, Y. Gupta, C. Hirata, J. Odegova, U.-L. Pen, J. B. Peterson, J. Roy, J. R. Shaw, K. Sigurdson, and T. Voytek. A simulation-calibrated limit on the H I power spectrum from the GMRT Epoch of Reionization experiment. MNRAS, 433:639–647, July 2013. doi: 10.1093/mnras/stt753.

[168] A. Parsons. The symmetric group in data permutation, with applications to high-bandwidth pipelined fft architectures. *Signal Processing Letters, IEEE*, 16 (6):477–480, 2009. ISSN 1070-9908. doi: 10.1109/LSP.2009.2016836.

[169] A. Parsons, D. Backer, C. Chang, D. Chapman, H. Chen, P. Crescini, C. de Jesus, C. Dick, P. Droz, D. MacMahon, K. Meder, J. Mock, V. Nagpal, B. Nikolic, A. Parsa, B. Richards, A. Siemion, J. Wawrzynek, D. Werthimer, and M. Wright. Petaop/second fpga signal processing for seti and radio astronomy. In *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, pages 2031–2035, 2006. doi: 10.1109/ACSSC.2006.355123.

[170] A. Parsons, J. Pober, M. McQuinn, D. Jacobs, and J. Aguirre. A Sensitivity and Array-configuration Study for Measuring the Power Spectrum of 21 cm Emission from Reionization. ApJ, 753:81, July 2012. doi: 10.1088/0004-637X/753/1/81.

[171] A. R. Parsons, D. C. Backer, G. S. Foster, M. C. H. Wright, R. F. Bradley, N. E. Gugliucci, C. R. Parashare, E. E. Benoit, J. E. Aguirre, D. C. Jacobs, C. L. Carilli, D. Herne, M. J. Lynch, J. R. Manley, and D. J. Werthimer. The

Precision Array for Probing the Epoch of Re-ionization: Eight Station Results. AJ, 139:1468–1480, April 2010. doi: 10.1088/0004-6256/139/4/1468.

[172] A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs, and D. F. Moore. A Per-baseline, Delay-spectrum Technique for Accessing the 21 cm Cosmic Reionization Signature. ApJ, 756:165, September 2012. doi: 10. 1088/0004-637X/756/2/165.

[173] A. R. Parsons, A. Liu, J. E. Aguirre, Z. S. Ali, R. F. Bradley, C. L. Carilli, D. R. DeBoer, M. R. Dexter, N. E. Gugliucci, D. C. Jacobs, P. Klima, D. H. E. MacMahon, J. R. Manley, D. F. Moore, J. C. Pober, I. I. Stefan, and W. P. Walbrugh. New Limits on 21 cm Epoch of Reionization from PAPER-32 Consistent with an X-Ray Heated Intergalactic Medium at z = 7.7. ApJ, 788:106, June 2014. doi: 10.1088/0004-637X/788/2/106.

[174] A. H. Patil, S. Zaroubi, E. Chapman, V. Jelić, G. Harker, F. B. Abdalla, K. M. B. Asad, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, S. Bus, B. Ciardi, S. Daiboo, E. R. Fernandez, A. Ghosh, H. Jensen, S. Kazemi, L. V. E. Koopmans, P. Labropoulos, M. Mevius, O. Martinez, G. Mellema, A. R. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, H. K. Vedantham, V. Veligatla, S. J. Wijnholds, and S. Yatawatta. Constraining the epoch of reionization with the variance statistic: simulations of the LOFAR case. MNRAS, 443:1113–1124, September 2014. doi: 10.1093/mnras/stu1178.

[175] U.-L. Pen. Fast power spectrum estimation. MNRAS, 346:619–626, December 2003. doi: 10.1046/j.1365-2966.2003.07118.x.

[176] N. Petrovic and S. P. Oh. Systematic effects of foreground removal in 21-cm surveys of reionization. MNRAS, 413:2103–2120, May 2011. doi: 10.1111/j. 1365-2966.2011.18276.x.

[177] B. Pindor, J. S. B. Wyithe, D. A. Mitchell, S. M. Ord, R. B. Wayth, and L. J. Greenhill. Subtraction of Bright Point Sources from Synthesis Images of the Epoch of Reionization. PASA, 28:46–57, January 2011. doi: 10.1071/AS10023.

[178] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. A&A, 571: A16, November 2014. doi: 10.1051/0004-6361/201321591.

[179] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. Planck 2015 results. XIII. Cosmological parameters. *ArXiv e-prints*, February 2015.

[180] J. Pober and others. Widefield Source Subtraction in 21 cm Epoch of Reionization Observations with the Murchison Widefield Array. *in preparation.*, 2015.

[181] J. C. Pober, A. R. Parsons, D. C. Jacobs, J. E. Aguirre, R. F. Bradley, C. L. Carilli, N. E. Gugliucci, D. F. Moore, and C. R. Parashare. A Technique for Primary Beam Calibration of Drift-scanning, Wide-field Antenna Elements. AJ, 143:53, February 2012. doi: 10.1088/0004-6256/143/2/53.

[182] J. C. Pober, A. R. Parsons, J. E. Aguirre, Z. Ali, R. F. Bradley, C. L. Carilli, D. DeBoer, M. Dexter, N. E. Gugliucci, D. C. Jacobs, P. J. Klima, D. MacMahon, J. Manley, D. F. Moore, I. I. Stefan, and W. P. Walbrugh. Opening the 21 cm Epoch of Reionization Window: Measurements of Foreground Isolation with PAPER. ApJ, 768:L36, May 2013. doi: 10.1088/2041-8205/768/2/L36.

[183] J. C. Pober, A. R. Parsons, D. R. DeBoer, P. McDonald, M. McQuinn, J. E. Aguirre, Z. Ali, R. F. Bradley, T.-C. Chang, and M. F. Morales. The Baryon Acoustic Oscillation Broadband and Broad-beam Array: Design Overview and Sensitivity Forecasts. AJ, 145:65, March 2013. doi: 10.1088/0004-6256/145/3/65.

[184] J. C. Pober, A. Liu, J. S. Dillon, J. E. Aguirre, J. D. Bowman, R. F. Bradley, C. L. Carilli, D. R. DeBoer, J. N. Hewitt, D. C. Jacobs, M. McQuinn, M. F. Morales, A. R. Parsons, M. Tegmark, and D. J. Werthimer. What Next-generation 21 cm Power Spectrum Measurements can Teach us About the Epoch of Reionization. ApJ, 782:66, February 2014. doi: 10.1088/0004-637X/782/2/66.

[185] J. C. Pober, Z. S. Ali, A. R. Parsons, M. McQuinn, J. E. Aguirre, G. Bernardi, R. F. Bradley, C. L. Carilli, C. Cheng, D. R. DeBoer, M. R. Dexter, S. R. Furlanetto, J. Grobbelaar, J. Horrell, D. C. Jacobs, P. J. Klima, S. A. Kohn, A. Liu, D. H. E. MacMahon, M. Maree, A. Mesinger, D. F. Moore, N. Razavi-Ghods, I. I. Stefan, W. P. Walbrugh, A. Walker, and H. Zheng. PAPER-64 Constraints On Reionization II: The Temperature Of The z=8.4 Intergalactic Medium. *ArXiv e-prints*, February 2015.

[186] J. Pritchard, K. Ichiki, A. Mesinger, R. B. Metcalf, A. Pourtsidou, M. Santos, F. Abdalla, T.-C. Chang, X. Chen, J. Weller, and S. Zaroubi. Cosmology from the EoR/Cosmic Dawn with the SKA. *ArXiv e-prints*, January 2015.

[187] J. R. Pritchard and S. R. Furlanetto. 21-cm fluctuations from inhomogeneous X-ray heating before reionization. MNRAS, 376:1680–1694, April 2007. doi: 10.1111/j.1365-2966.2007.11519.x.

[188] J. R. Pritchard and A. Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, August 2012. doi: 10.1088/0034-4885/75/8/086901.

[189] U. Rau, S. Bhatnagar, M. A. Voronkov, and T. J. Cornwell. Advances in Calibration and Imaging Techniques in Radio Interferometry. *IEEE Proceedings*, 97:1472–1481, August 2009. doi: 10.1109/JPROC.2009.2014853.

[190] B. E. Robertson, R. S. Ellis, S. R. Furlanetto, and J. S. Dunlop. Cosmic Reionization and Early Star-Forming Galaxies: A Joint Analysis of New Constraints from Planck and Hubble Space Telescope. *ArXiv e-prints*, February 2015.

[191] A. E. E. Rogers and J. D. Bowman. Spectral Index of the Diffuse Radio Background Measured from 100 to 200 MHz. AJ, 136:641–648, August 2008. doi: 10.1088/0004-6256/136/2/641.

[192] S. Salvini and S. J. Wijnholds. Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications. A&A, 571: A97, November 2014. doi: 10.1051/0004-6361/201424487.

[193] N. J. Sanchez. On the instrumentation of the Omniscope. Master's thesis, Massachusetts Institute of Technology, 2011. URL `http://hdl.handle.net/1721.1/66807`.

[194] M. G. Santos and A. Cooray. Cosmological and astrophysical parameter measurements with 21-cm anisotropies during the era of reionization. Phys. Rev. D, 74(8):083517–+, October 2006. doi: 10.1103/PhysRevD.74.083517.

[195] David Schramm. Shadows of creation: The dark matter of the universe. *Bulletin of the American Academy of Arts and Sciences*, 50(6):pp. 40–51, 1997. ISSN 0002712X. URL `http://www.jstor.org/stable/3824359`.

[196] F. R. Schwab. Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. AJ, 89:1076–1081, July 1984. doi: 10.1086/113605.

[197] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins, and M. Sitwell. All-Sky Interferometry with Spherical Harmonic Transit Telescopes. *ArXiv e-prints*, February 2013.

[198] J. R. Shaw, K. Sigurdson, M. Sitwell, A. Stebbins, and U.-L. Pen. Coaxing Cosmic 21cm Fluctuations from the Polarized Sky using m-mode Analysis. *ArXiv e-prints*, January 2014.

[199] J. R Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. *unpublished, available online*, August 1994. URL `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`.

[200] O. B. Slee. Radio sources observed with the Culgoora circular array. *Australian Journal of Physics*, 48:143–186, 1995.

[201] O. M. Smirnov. Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism. A&A, 527:A106, March 2011. doi: 10.1051/0004-6361/201016082.

[202] A. Songaila and L. L. Cowie. The Evolution of Lyman Limit Absorption Systems to Redshift Six. ApJ, 721:1448–1466, October 2010. doi: 10.1088/0004-637X/721/2/1448.

[203] I. S. Sullivan, M. F. Morales, B. J. Hazelton, W. Arcus, D. Barnes, G. Bernardi, F. H. Briggs, J. D. Bowman, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. deSouza, D. Emrich, B. M. Gaensler, R. Goeke, L. J. Greenhill, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, E. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya Shankar, K. S. Srivani, J. Stevens, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. Fast Holographic Deconvolution: A New Technique for Precision Radio Interferometry. ApJ, 759:17, November 2012. doi: 10.1088/0004-637X/759/1/17.

[204] A. Sutinjo, J. O'Sullivan, E. Lenc, R. B. Wayth, S. Padhi, P. Hall, and S. J. Tingay. Understanding Instrumental Stokes Leakage in Murchison Widefield Array Polarimetry. ArXiv e-prints, December 2014.

[205] P. M. Sutter, B. D. Wandelt, and S. S. Malu. Bayesian Angular Power Spectrum Analysis of Interferometric Data. ApJS, 202:9, September 2012. doi: 10.1088/0067-0049/202/1/9.

[206] P. M. Sutter, B. D. Wandelt, J. D. McEwen, E. F. Bunn, A. Karakci, A. Korotkov, P. Timbie, G. S. Tucker, and L. Zhang. Probabilistic image reconstruction for radio interferometers. MNRAS, 438:768–778, February 2014. doi: 10.1093/mnras/stt2244.

[207] E. R. Switzer and A. Liu. Erasing the Variable: Empirical Foreground Discovery for Global 21 cm Spectrum Experiments. ApJ, 793:102, October 2014. doi: 10.1088/0004-637X/793/2/102.

[208] M. Tegmark. How to measure CMB power spectra without losing information. Phys. Rev. D, 55:5895–5907, May 1997. doi: 10.1103/PhysRevD.55.5895.

[209] M. Tegmark. How to Make Maps from Cosmic Microwave Background Data without Losing Information. ApJ, 480:L87, May 1997. doi: 10.1086/310631.

[210] M. Tegmark and M. Zaldarriaga. Fast Fourier transform telescope. Phys. Rev. D, 79(8):083530, April 2009. doi: 10.1103/PhysRevD.79.083530.

[211] M. Tegmark and M. Zaldarriaga. Omniscopes: Large area telescope arrays with only N log N computational cost. Phys. Rev. D, 82(10):103501, November 2010. doi: 10.1103/PhysRevD.82.103501.

[212] M. Tegmark, A. N. Taylor, and A. F. Heavens. Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets? ApJ, 480: 22–35, May 1997.

[213] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay. Measuring the Galaxy Power Spectrum with Future Redshift Surveys. *ApJ*, May 1998.

[214] M. Tegmark, A. J. S. Hamilton, and Y. Xu. The power spectrum of galaxies in the 2dF 100k redshift survey. MNRAS, 335:887–908, October 2002. doi: 10.1046/j.1365-8711.2002.05622.x.

[215] M. Tegmark, M. R. Blanton, M. A. Strauss, and et al. The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, 606: 702–740, May 2004. doi: 10.1086/382125.

[216] A Richard Thompson, James M Moran, and George W Swenson. *Interferometry and Synthesis in Radio Astronomy; 2nd ed.* Wiley-VCH, Weinheim, 2001.

[217] N. Thyagarajan and others. Confirmation of Wide-Field Signatures in Redshifted 21 cm Power Spectra Using Murchison Widefield Array Observations. *in preparation.*, 2015.

[218] N. Thyagarajan, N. Udaya Shankar, R. Subrahmanyan, W. Arcus, G. Bernardi, J. D. Bowman, F. Briggs, J. D. Bunton, R. J. Cappallo, B. E. Corey, L. deSouza, D. Emrich, B. M. Gaensler, R. F. Goeke, L. J. Greenhill, B. J. Hazelton, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. H. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, R. A. Remillard, A. E. E. Rogers, D. Anish Roshi, J. E. Salah, R. J. Sault, K. S. Srivani, J. B. Stevens, P. Thiagaraj, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. J. Williams, C. L. Williams, and J. S. B. Wyithe. A Study of Fundamental Limitations to Statistical Detection of Redshifted H I from the Epoch of Reionization. ApJ, 776:6, October 2013. doi: 10.1088/0004-637X/776/1/6.

[219] N. Thyagarajan, D. C. Jacobs, J. D. Bowman, N. Barry, A. P. Beardsley, G. Bernardi, F. Briggs, R. J. Cappallo, P. Carroll, B. E. Corey, A. de Oliveira-Costa, J. S. Dillon, D. Emrich, A. Ewall-Wice, L. Feng, R. Goeke, L. J. Greenhill, B. J. Hazelton, J. N. Hewitt, N. Hurley-Walker, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, H.-S. Kim, P. Kittiwisit, E. Kratzenberg, E. Lenc, J. Line, A. Loeb, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, A. R. Neben, D. Oberoi, A. R. Offringa, S. M. Ord, S. Paul, B. Pindor, J. C. Pober, T. Prabu, P. Procopio, J. Riding, A. E. E. Rogers, A. Roshi, N. Udaya Shankar, S. K. Sethi, K. S. Srivani, R. Subrahmanyan, I. S. Sullivan, M. Tegmark, S. J. Tingay, C. M. Trott, M. Waterson, R. B. Wayth, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams,

C. Wu, and J. S. B. Wyithe. Foregrounds in Wide-Field Redshifted 21 cm Power Spectra. *ArXiv e-prints*, February 2015.

[220] S. J. Tingay, R. Goeke, J. D. Bowman, D. Emrich, S. M. Ord, D. A. Mitchell, M. F. Morales, T. Booler, B. Crosse, R. B. Wayth, C. J. Lonsdale, S. Tremblay, D. Pallot, T. Colegate, A. Wicenec, N. Kudryavtseva, W. Arcus, D. Barnes, G. Bernardi, F. Briggs, S. Burns, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. Desouza, B. M. Gaensler, L. J. Greenhill, P. J. Hall, B. J. Hazelton, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, M. J. Lynch, B. Mckinley, S. R. Mcwhirter, E. Morgan, D. Oberoi, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya-Shankar, F. Schlagenhaufer, K. S. Srivani, J. Stevens, R. Subrahmanyan, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. PASA, 30:e007, June 2013. doi: 10.1017/pasa.2012.007.

[221] P. Tozzi, P. Madau, A. Meiksin, and M. J. Rees. The first light seen in the redshifted 21-cm radiation. *Nuclear Physics B Proceedings Supplements*, 80: C509+, January 2000.

[222] P. Tozzi, P. Madau, A. Meiksin, and M. J. Rees. Radio Signatures of H I at High Redshift: Mapping the End of the "Dark Ages". *ApJ*, 528:597–606, January 2000. doi: 10.1086/308196.

[223] H. Trac and R. Cen. Radiative Transfer Simulations of Cosmic Reionization. I. Methodology and Initial Results. ApJ, 671:1–13, December 2007. doi: 10. 1086/522566.

[224] C. M. Trott. Comparison of Observing Modes for Statistical Estimation of the 21 cm Signal from the Epoch of Reionisation. PASA, 31:e026, July 2014. doi: 10.1017/pasa.2014.23.

[225] C. M. Trott, R. B. Wayth, and S. J. Tingay. The Impact of Point-source Subtraction Residuals on 21 cm Epoch of Reionization Estimation. ApJ, 757: 101, September 2012. doi: 10.1088/0004-637X/757/1/101.

[226] D. Tseliakhovich and C. Hirata. Relative velocity of dark matter and baryonic fluids and the formation of the first structures. Phys. Rev. D, 82(8):083520, October 2010. doi: 10.1103/PhysRevD.82.083520.

[227] M. Valdés, C. Evoli, A. Mesinger, A. Ferrara, and N. Yoshida. The nature of dark matter from the global high-redshift H I 21 cm signal. MNRAS, 429: 1705–1716, February 2013. doi: 10.1093/mnras/sts458.

[228] S. van der Tol, B. D. Jeffs, and A.-J. . van der Veen. Self-Calibration for the LOFAR Radio Astronomical Array. *IEEE Transactions on Signal Processing*, 55:4497–4510, September 2007. doi: 10.1109/TSP.2007.896243.

[229] M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A.-J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R.-J. Dettmar, G. van Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel, J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J.-M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J.-P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H.-J. Stiepel, K. Stuurwold, M. Tagger, Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A.-J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans, P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus, and J. van Zwieten. LOFAR: The LOw-Frequency ARray. A&A, 556:A2, August 2013. doi: 10.1051/0004-6361/201220873.

[230] H. Vedantham, N. Udaya Shankar, and R. Subrahmanyan. Imaging the Epoch of Reionization: Limitations from Foreground Confusion and Imaging Algorithms. ApJ, 745:176, February 2012. doi: 10.1088/0004-637X/745/2/176.

[231] X. Wang, M. Tegmark, M. G. Santos, and L. Knox. 21 cm Tomography with Foregrounds. *ApJ*, 650:529–537, October 2006. doi: 10.1086/506597.

[232] N Wiener. The interpolation, extrapolation and smoothing of stationary time series. Report of the Services 19, Massachusetts Institute of Technology, February 1942. Research Project DIC-6037.

[233] M. H. Wieringa. An investigation of the telescope based calibration methods 'redundancy' and 'self-cal'. *Experimental Astronomy*, 2:203–225, 1992. doi: 10.1007/BF00420576.

[234] C. L. Williams, J. N. Hewitt, A. M. Levine, A. de Oliveira-Costa, J. D. Bowman, F. H. Briggs, B. M. Gaensler, L. L. Hernquist, D. A. Mitchell, M. F. Morales, S. K. Sethi, R. Subrahmanyan, E. M. Sadler, W. Arcus, D. G. Barnes, G. Bernardi, J. D. Bunton, R. C. Cappallo, B. W. Crosse, B. E. Corey, A. Deshpande, L. deSouza, D. Emrich, R. F. Goeke, L. J. Greenhill, B. J. Hazelton, D. Herne, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, E. H. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, D. Anish Roshi, J. E. Salah, R. J. Sault, N. Udaya Shankar, K. S. Srivani, J. B. Stevens, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. J. Williams, and J. S. B. Wyithe. Low-frequency Imaging of Fields at High Galactic Latitude with the Murchison Widefield Array 32 Element Prototype. ApJ, 755:47, August 2012. doi: 10.1088/0004-637X/755/1/47.

[235] C. J. Willott, S. Rawlings, K. M. Blundell, M. Lacy, and S. A. Eales. The radio luminosity function from the low-frequency 3CRR, 6CE and 7CRS complete samples. MNRAS, 322:536–552, April 2001. doi: 10.1046/j.1365-8711.2001.04101.x.

[236] R. J. Wilman, L. Miller, M. J. Jarvis, T. Mauch, F. Levrier, F. B. Abdalla, S. Rawlings, H.-R. Klöckner, D. Obreschkow, D. Olteanu, and S. Young. A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes. MNRAS, 388:1335–1348, August 2008. doi: 10.1111/j.1365-2966.2008.13486.x.

[237] T. A. Wood and G. Chan. Simulation of Stationary Gaussian Processes in [0,1]d. *Journal of Computational and Graphical Statistics*, 3(4):409–432, December 1994.

[238] S. A. Wouthuysen. On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line. AJ, 57:31–32, 1952. doi: 10.1086/106661.

[239] X. Wu. Probing the Epoch of Reionization with 21CMA: Status and Prospects. In *American Astronomical Society Meeting Abstracts #213*, volume 41 of *Bulletin of the American Astronomical Society*, page 226.05, January 2009.

[240] J. S. B. Wyithe and A. Loeb. Self-regulated Growth of Supermassive Black Holes in Galaxies as the Origin of the Optical and X-Ray Luminosity Functions of Quasars. ApJ, 595:614–623, October 2003. doi: 10.1086/377475.

[241] J. S. B. Wyithe, A. Loeb, and P. M. Geil. Baryonic acoustic oscillations in 21-cm emission: a probe of dark energy out to high redshifts. *MNRAS*, 383: 1195–1209, January 2008. doi: 10.1111/j.1365-2966.2007.12631.x.

[242] S. Wyithe, P. M. Geil, and H. Kim. Imaging HII Regions from Galaxies and Quasars During Reionisation with SKA. *ArXiv e-prints*, January 2015.

[243] Y. Xu, X. Chen, Z. Fan, H. Trac, and R. Cen. The 21 cm Forest as a Probe of the Reionization and The Temperature of the Intergalactic Medium. ApJ, 704: 1396–1404, October 2009. doi: 10.1088/0004-637X/704/2/1396.

[244] S. Yatawatta, A. G. de Bruyn, M. A. Brentjens, P. Labropoulos, V. N. Pandey, S. Kazemi, S. Zaroubi, L. V. E. Koopmans, A. R. Offringa, V. Jelić, O. Martinez Rubi, V. Veligatla, S. J. Wijnholds, W. N. Brouw, G. Bernardi, B. Ciardi, S. Daiboo, G. Harker, G. Mellema, J. Schaye, R. Thomas, H. Vedantham, E. Chapman, F. B. Abdalla, A. Alexov, J. Anderson, I. M. Avruch, F. Batejat, M. E. Bell, M. R. Bell, M. Bentum, P. Best, A. Bonafede, J. Bregman, F. Breitling, R. H. van de Brink, J. W. Broderick, M. Brüggen, J. Conway, F. de Gasperin, E. de Geus, S. Duscha, H. Falcke, R. A. Fallows, C. Ferrari, W. Frieswijk, M. A. Garrett, J. M. Griessmeier, A. W. Gunst, T. E. Hassall, J. W. T. Hessels, M. Hoeft, M. Iacobelli, E. Juette, A. Karastergiou, V. I. Kondratiev, M. Kramer, M. Kuniyoshi, G. Kuper, J. van Leeuwen, P. Maat, G. Mann, J. P. McKean, M. Mevius, J. D. Mol, H. Munk, R. Nijboer, J. E. Noordam, M. J. Norden, E. Orru, H. Paas, M. Pandey-Pommier, R. Pizzo, A. G. Polatidis, W. Reich, H. J. A. Röttgering, J. Sluman, O. Smirnov, B. Stappers, M. Steinmetz, M. Tagger, Y. Tang, C. Tasse, S. ter Veen, R. Vermeulen, R. J. van Weeren, M. Wise, O. Wucknitz, and P. Zarka. Initial deep LOFAR observations of epoch of reionization windows. I. The north celestial pole. A&A, 550: A136, February 2013. doi: 10.1051/0004-6361/201220874.

[245] O. Zahn, A. Mesinger, M. McQuinn, H. Trac, R. Cen, and L. E. Hernquist. Comparison of reionization models: radiative transfer simulations and approximate, seminumeric models. MNRAS, 414:727–738, June 2011. doi: 10.1111/j.1365-2966.2011.18439.x.

[246] M. Zaldarriaga, S. R. Furlanetto, and L. Hernquist. 21 Centimeter Fluctuations from Cosmic Gas at High Redshifts. *ApJ*, 608:622–635, June 2004. doi: 10. 1086/386327.

[247] H. Zheng, M. Tegmark, V. Buza, J. S. Dillon, H. Gharibyan, J. Hickish, E. Kunz, A. Liu, J. Losh, A. Lutomirski, S. Morrison, S. Narayanan, A. Perko, D. Rosner, N. Sanchez, K. Schutz, S. M. Tribiano, M. Zaldarriaga, K. Zarb Adami, I. Zelko, K. Zheng, R. Armstrong, R. F. Bradley, M. R.

Dexter, A. Ewall-Wice, A. Magro, M. Matejek, E. Morgan, A. R. Neben, Q. Pan, C. M. Peterson, M. Su, J. Villasenor, C. L. Williams, H.-I Yang, and Y. Zhu. Mapping our universe in 3d with miteor. *2013 IEEE International Symposium on Phased Array Systems and Technology*, Oct 2013. URL `http://dx.doi.org/10.1109/ARRAY.2013.6731928`.

[248] H. Zheng, M. Tegmark, V. Buza, J. S. Dillon, H. Gharibyan, J. Hickish, E. Kunz, A. Liu, J. Losh, A. Lutomirski, S. Morrison, S. Narayanan, A. Perko, D. Rosner, N. Sanchez, K. Schutz, S. M. Tribiano, M. Valdez, H. Yang, K. Z. Adami, I. Zelko, K. Zheng, R. P. Armstrong, R. F. Bradley, M. R. Dexter, A. Ewall-Wice, A. Magro, M. Matejek, E. Morgan, A. R. Neben, Q. Pan, R. F. Penna, C. M. Peterson, M. Su, J. Villasenor, C. L. Williams, and Y. Zhu. MITEoR: a scalable interferometer for precision 21 cm cosmology. MNRAS, 445:1084–1103, December 2014. doi: 10.1093/mnras/stu1773.